

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 26, 2005, 12:53:47 ; Search time 8121.42 Seconds
(without alignment)
10694.773 Million cell updates/sec

Title: US-09-555-529-1
Perfect score: 1528
Sequence: 1 ctgaattcagccgcctg.....atgtattataaaaaaaa 1528

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2006000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_in.*
- 3: gb_env.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pr.*
- 9: gb_ro.*
- 10: gb_sts.*
- 11: gb_sv.*
- 12: gb_un.*
- 13: gb_vl.*
- 14: gb_htg.*
- 15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1528	100.0	1528	6	AX003308 Sequence
2	1528	100.0	1528	8	HSJ5273
3	1453	95.1	2069	8	BC017309 Homo sapi
4	1002	65.6	1002	6	AX003311 Sequence
5	999.8	65.4	1414	9	MMKIN17
6	996.4	65.2	1386	9	BC028860 Mus muscu
7	990.6	64.8	1390	6	AX003331 Sequence
8	969.2	63.4	221547	14	AC119716
9	955.2	62.5	1117	6	CQ722054 Sequence
10	759	49.7	130262	14	AC166154 Mus muscu
11	759	49.7	140215	9	AC140205 Mus muscu
12	705	46.1	1483	5	BC074606 Xenopus t
13	685	44.8	144118	14	AC141995 Rattus no
14	673.2	44.1	180882	14	AC114615 Mus muscu
15	673.2	44.1	182400	14	AC124010 Mus muscu
16	673.2	44.1	185378	9	AL772367 Mouse DNA
17	672.4	44.0	1501	5	BC079988 Xenopus l
18	663.8	43.4	2710	9	BC058169 Mus muscu

19	653.4	42.8	1002	6	AX003310 Sequence
20	475	31.1	1102	6	AX003309 Sequence
21	465.4	30.5	175480	14	AC127074 Rattus no
22	465.4	30.5	247615	14	AC098061 Rattus no
23	436.4	28.6	461	6	CQ684251 Sequence
24	402.8	26.4	177276	14	AC154247 Mus muscu
25	380.6	24.9	1372	2	AK113340 Ciona int
26	337	22.1	174361	8	AL158044 Human DNA
27	337	22.1	194563	14	AL391687 Homo sapi
28	333	21.8	333	6	AX003340 Sequence
29	300.2	19.6	1467	2	AY084184 Drosophill
30	298.6	19.5	1241	6	CQ613281 Sequence
31	284	18.6	300	6	BD220437 Human gen
32	261.6	17.1	333	6	AX003341 Sequence
33	249.6	16.3	616	10	BV448133 KIN 9291
34	247.2	16.2	3313	6	CQ613280 Sequence
35	247.2	16.2	31807	14	AC017679 Drosophill
36	247.2	16.2	185918	2	AC009839 Drosophill
37	247.2	16.2	301136	2	AB003591 Drosophill
38	247	16.2	84798	14	AF007827 Lotus cor
39	224.8	14.7	1563	15	AF360132 Arabidops
40	224.8	14.7	87967	15	AC005223 Arabidops
41	224.8	14.7	1267	15	AY051011 Arabidops
42	218	14.3	760	2	DME6528 Arabidops
43	209.4	13.7	4394	6	CQ613259 Drosophill
44	209.4	13.7	4582	6	CQ596774 Sequence
45	208.2	13.6	1824	15	AK069396 Oryza sat

ALIGNMENTS

RESULT 1
AX003308 LOCUS
DEFINITION Sequence 1 from Patent WO929845.
ACCESSION AX003308
VERSION AX003308.1 GI:9927125
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1
AUTHORS Angulo-Mora, J.F. and Mauffrey, P.
TITLES Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 1 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)

FEATURES
source Location/Qualifiers
1..1528
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 1528; DB 6; Length 1528;
Best Local Similarity 100.0%; Pred. No. 2e-311;
Matches 1528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTAGATTTCAGCGCGCGCTGTAATTCCTAGAACTGGGTCAGAAAGTGATCGTCCGCTGG 60
DB 1 CTAGATTTCAGCGCGCGCTGTAATTCCTAGAACTGGGTCAGAAAGTGATCGTCCGCTGG 60
QY 61 TCGCCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGATCAAGTCCA 120
DB 61 TCGCCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGATCAAGTCCA 120
QY 121 AGGGGCTGCAGAAAGCTACGCTGGTATTCGAGATGTCGAGAGAGCTCCGGGACGAGA 180
DB 121 AGGGGCTGCAGAAAGCTACGCTGGTATTCGAGATGTCGAGAGAGCTCCGGGACGAGA 180

Bouffard G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
 Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
 Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.E.,
 Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2069)

Strausberg, R.

Direct Submission

Submitted (05-NOV-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

On Dec 19, 2003 this sequence version replaced gi:16878214.

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@hgrl.nih.gov

Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,

Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,

Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,

Maduro, Q.B., Masiello, C., Maskeri, B., Mastriani, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,

Teurigon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IHAL Plate: 39 Row: 0 Column: 18

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 13124882.

FEATURES

source

1..2069

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:29595 IMAGE:5089578"

/tissue_type="Pancreas, epithelioid carcinoma"

/clone_lib="NIH MGC_42"

/lab_host="DH10B-R"

/note="vector: pOTB7"

1..2069

/gene="KIN"

/note="synonyms: KIN17, BTCD"

/db_xref="GeneID:22944"

/db_xref="MIM:601720"

12..1193

/gene="KIN"

/codon_start=1

/product="HsKIN17 protein"

/protein_id="AAH17309.1"

/db_xref="GI:16878215"

/db_xref="GeneID:22944"

/db_xref="MIM:601720"

/translation="WGKSDFLTPKAIANRIKSLQKLRLWYCOMQCRDENGFKCH

CMSSHQQLLASENPQFMDYSEPRNDFLELLRRFCGTRKHVNNIYVNEIYISHR

EHIHNAQTOWETLDFPKWLGREGCKVDQETPKWYLOYIDRDETIRROLELEKKK

QDLDEKTAKEIQVNRGLGKEQVPTFELSRENDEKVTFNLSKGCSSGAT

SSKSTLGSALKTIGSSASVKRKESSQTSQSKKKKKSALDEIMEIEBKRTAR

TDYWLQPEIIVKIIITKKLGEYHKKAIVKEIDKYTAVMKMDSGKLDQDTHLET

VIPAPGKRILVNGYRGNEGTLESINEKTFESATIVETGPLKGRRVEGIQYEDISKL

"A"

,ORIGIN

Query Match	95.1%	Score 1453;	DB 8;	Length 2069;
Best Local Similarity	99.7%	Pred. No. 1.2e-295;		
Matches 1467;	Conservative 0;	Mismatches 4;	Gaps 1;	
QY	55	CCGTGGTCCGATCGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAAGGATCA	114	
Db	1	CCGTGGTCCGATCGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAAGGATCA	60	
QY	115	AGTCCAAAGGGGCTGCAGAAGCTACGCTGGTATTTGCCAGATGTGCCAGAAGCAGTCCGGG	174	
Db	61	AGTCCAAAGGGGCTGCAGAAGCTACGCTGGTATTTGCCAGATGTGCCAGAAGCAGTCCGGG	120	
QY	175	ACGAGAATGGCTTTAAGTGTCAATTTGATGCCGAATCTCATCAGAGACAATATTGCTGG	234	
Db	121	ACGAGAATGGCTTTAAGTGTCAATTTGATGCCGAATCTCATCAGAGACAATATTGCTGG	180	
QY	235	CTTCAGAAAATCCTCAGCAGTTTATGATTTATTTTCAGAGGAATTCGGAATGACTTTC	294	
Db	181	CTTCAGAAAATCCTCAGCAGTTTATGATTTATTTTCAGAGGAATTCGGAATGACTTTC	240	
QY	295	TAGAACTTCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAAACAATTTGTTCTACAACG	354	
Db	241	TAGAACTTCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAAACAATTTGTTCTACAACG	300	
QY	355	AATACATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAATCTGACTG	414	
Db	301	AATACATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAATCTGACTG	360	
QY	415	ATTTTACTAGTGGCTCGGAGAGAGAGCTTTGTCAAAGTGGAGAGACACACAAAGGCT	474	
Db	361	ATTTTACTAGTGGCTCGGAGAGAGAGCTTTGTCAAAGTGGAGAGACACACAAAGGCT	420	
QY	475	GGTATATTTCAGTACATAGACAGGACCCAGAAACTATCCCGCGGCAATCGAACTGGAGA	534	
Db	421	GGTATATTTCAGTACATAGACAGGACCCAGAAACTATCCCGCGGCAATCGAACTGGAGA	480	
QY	535	AAAAGAAAAGCAGGACCTTGATGATGAAGAAAATTCGCAATTTATTGAAGAGCAAG	594	
Db	481	AAAAGAAAAGCAGGACCTTGATGATGAAGAAAATTCGCAATTTATTGAAGAGCAAG	540	
QY	595	TGAGAAGAGCCTTGAAGGAAGGAAGCAGGAGTCCCTACTTTTACGGAATTAAGCAGAG	654	
Db	541	TGAGAAGAGCCTTGAAGGAAGGAAGCAGGAGTCCCTACTTTTACGGAATTAAGCAGAG	600	
QY	655	AAAATGATGAAGAGAAAAGTCACGTTTAAATTTGAGTAAAGGAGCATGTAGCTCATCCGGAG	714	
Db	601	AAAATGATGAAGAGAAAAGTCACGTTTAAATTTGAGTAAAGGAGCATGTAGCTCATCCGGAG	660	
QY	715	CAACATCTTCAGTCAAGTACTCTGGGACCGAGTGCATCGAAGACATAGGAAGTTTCAG	774	
Db	661	CAACATCTTCAGTCAAGTACTCTGGGACCGAGTGCATCGAAGACATAGGAAGTTTCAG	720	
QY	775	CATCAGTGAAGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAAGAAAGAAA	834	
Db	721	CATCAGTGAAGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAAGAAAGAAA	780	
QY	835	AGAAATCTGCACTGGATGAAATCATGGAGATTGAAGAGGAAAAGAAAAGAAATCTCCCGAA	894	
Db	781	AGAAATCTGCACTGGATGAAATCATGGAGATTGAAGAGGAAAAGAAAAGAAATCTCCCGAA	840	
QY	895	CAGACTACTGGCTACAGCTGAAATTTATTGTGAAATTTATAACCAAGAACTGGAGAGA	954	
Db	841	CAGACTACTGGCTACAGCTGAAATTTATTGTGAAATTTATAACCAAGAACTGGAGAGA	900	
QY	955	AATATATAAGAAAAGGCTATTGTTAAGGAAGTAAATTTGACAAATATACAGCTGTGTGA	1014	
Db	901	AATATATAAGAAAAGGCTATTGTTAAGGAAGTAAATTTGACAAATATACAGCTGTGTGA	960	
QY	1015	AGATGATTGATTCTGGAGACAAGCTGAAATTTGACCAAGCTTATTTAGAGACAGTAATTC	1074	
Db	961	AGATGATTGATTCTGGAGACAAGCTGAAATTTGACCAAGCTTATTTAGAGACAGTAATTC	1020	

QY	1075	CAGCACCAGGAAAGAAATCTAGTTTAAATGGAGGCTACAGAGAAATGAAGGTACCC	1134
DB	1021	CAGCACCAGGAAAGAAATCTAGTTTAAATGGAGGCTACAGAGAAATGAAGGTACCC	1080
QY	1135	TAGAAATCCATCAATCAGAGAGACTTTTTCAGCTACTATCGTCAATGAAACTGGCCCTTTAA	1194
DB	1081	TAGAAATCCATCAATCAGAGAGACTTTTTCAGCTACTATCGTCAATGAAACTGGCCCTTTAA	1140
QY	1195	AAGGACGAGAGTGAAGGAATCAATATGAAGACATTTCTAAATCTGCCTGAGTTGAA	1254
DB	1141	AAGGACGAGAGTGAAGGAATCAATATGAAGACATTTCTAAATCTGCCTGAGTTGAA	1200
QY	1255	AAATTTGTTAAACAATACATTTAAATCTTAAAGCATCAAAATTTGGTTCGCAAGGCATTA	1314
DB	1201	AAATTTGTTAAACAATACATTTAAATCTTAAAGCATCAAAATTTGGTTCGCAAGGCATTA	1260
QY	1315	GAGACTCTACTGTTAGGAGTATATCTTTTGTATATAAAACAAACAGGTTTTTGAATAAT	1374
DB	1261	GAGACTCTACTGTTAGGAGTATATCTTTTGTATATAAAACAAACAGGTTTTTGAATAAT	1320
QY	1375	TACTGTA-----TAGTTGTTTCAGCTAACTTTGAGAGAAATTTAATATGTCATGAGT	1430
DB	1321	TACTGTAATAGTTAGTTGTTTCAGCTAACTTTGAGAGAAATTTAATATGTCATGAGT	1380
QY	1431	ATCAAACTATGTAATTTGTCCTGCTTATTTTGTATTTGTTTCTTGTAAATTTACTTTGATG	1490
DB	1381	ATCAAACTATGTAATTTGTCCTGCTTATTTTGTATTTGTTTCTTGTAAATTTACTTTGATG	1440
QY	1491	TTATATCTTTCAATTAAGAAATGTTATTATATAA	1521
DB	1441	TTATATCTTTCAATTAAGAAATGTTATTATATAA	1471
RESULT 4			
AX003311			
LOCUS	AX003311	1002 bp	DNA linear PAT 24-AUG-2000
DEFINITION	Sequence 4 from Patent WO9929845.		
ACCESSION	AX003311		
VERSION	AX003311.1	GI:9927128	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1		
AUTHORS	Angulo-Mora, J.F. and Mauffrey, P.		
TITLE	Sequences coding for kin17 protein and their applications		
JOURNAL	Patent: WO 9929845-A 4 17-JUN-1999;		
	ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE		
	(FR)		
FEATURES	Location/Qualifiers		
source	1..1002		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	65.6%;	Score 1002;	DB 6; Length 1002;
Best Local Similarity	100.0%;	Pred. No. 1.4e-200;	
Matches 1002;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	215	TCAGAGACAACATATGCTGGCTTCAGAAATCTCAGCAGTATGATATTTTTCAGA	274
DB	1	TCAGAGACAACATATGCTGGCTTCAGAAATCTCAGCAGTATGATATTTTTCAGA	60
QY	275	GGAAATCCGAATGACATTTCTAGAACTTCTCAGAGAGCGTTTGGCACTAAAGGTTCCA	334
DB	61	GGAAATCCGAATGACATTTCTAGAACTTCTCAGAGAGCGTTTGGCACTAAAGGTTCCA	120
QY	335	CAACAACATTTGCTACAAACAAATACATCAGCCACCGAGAGACATCCACATGAATGCCAC	394
DB	121	CAACAACATTTGCTACAAACAAATACATCAGCCACCGAGAGACATCCACATGAATGCCAC	180

QY	395	TCAGTGGGAAACTCTGACTGATTTTACTAAGTGGCTGGGACAGAAAGCTTTGTGCAAAAGT	454
DB	181	TCAGTGGGAAACTCTGACTGATTTTACTAAGTGGCTGGGACAGAAAGCTTTGTGCAAAAGT	240
QY	455	GGACGAGACACAAAGGCTGTTATATTCAGTATACATAGACAGGACCCAGAAATCTATCCG	514
DB	241	GGACGAGACACAAAGGCTGTTATATTCAGTATACATAGACAGGACCCAGAAATCTATCCG	300
QY	515	CGGCACTCGAATCGGAGAGAAAGAAAAGCAGGACCTTGATGATGAAGAAAACCTGTC	574
DB	301	CGGCACTCGAATCGGAGAGAAAGAAAAGCAGGACCTTGATGATGAAGAAAACCTGTC	360
QY	575	CAAAATTTATGAAGAGCAAGTGAAGAGGCTGGAAGGGAAGGAACAGGAGTCCCTAC	634
DB	361	CAAAATTTATGAAGAGCAAGTGAAGAGGCTGGAAGGGAAGGAACAGGAGTCCCTAC	420
QY	635	TTTTCACGAAATTAAGCAGAGAAAATGATGAAGAGAAAGTACGTTTAAATTTGAGTAAAGG	694
DB	421	TTTTCACGAAATTAAGCAGAGAAAATGATGAAGAGAAAGTACGTTTAAATTTGAGTAAAGG	480
QY	695	AGCATGTAGCTCATCCGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCCT	754
DB	481	AGCATGTAGCTCATCCGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCCT	540
QY	755	GAAGACGATAGGAAGTTCAGCATCAGTGAACCAAGAAATCTTCCAGAGCTCAAGTACTCTGGGACCGAGTGCCT	814
DB	541	GAAGACGATAGGAAGTTCAGCATCAGTGAACCAAGAAATCTTCCAGAGCTCAAGTACTCTGGGACCGAGTGCCT	600
QY	815	GTCTAAAGAAAAGAAAAGAAAATCTGCATCAGTGAACCAAGAAATCTTCCAGAGCTCAAGTACTCTGGGACCGAGTGCCT	874
DB	601	GTCTAAAGAAAAGAAAAGAAAATCTGCATCAGTGAACCAAGAAATCTTCCAGAGCTCAAGTACTCTGGGACCGAGTGCCT	660
QY	875	AAAGAAAAGAACTGCCGACAGCTACTGCTCAGAGCTGAAATCTTGTGCAAAATAT	934
DB	661	AAAGAAAAGAACTGCCGACAGCTACTGCTCAGAGCTGAAATCTTGTGCAAAATAT	720
QY	935	AACCAAGAACTGGGAGAGAAATATCATAAGAAAAGGCTATTGTTAAGGAAGTAAATGA	994
DB	721	AACCAAGAACTGGGAGAGAAATATCATAAGAAAAGGCTATTGTTAAGGAAGTAAATGA	780
QY	995	CAAAATATACAGCTGTTGGAAGATGATGATCTCGAGACAAAGCTGAAACTTGACACAGAC	1054
DB	781	CAAAATATACAGCTGTTGGAAGATGATGATCTCGAGACAAAGCTGAAACTTGACACAGAC	840
QY	1055	TCATTTAGACAGTAATTCAGACCCAGGAAAGAAATCTAGTTTAAATGGAGGCTA	1114
DB	841	TCATTTAGACAGTAATTCAGACCCAGGAAAGAAATCTAGTTTAAATGGAGGCTA	900
QY	1115	CAGAGAAATGAAGGTACCTAGAAATCCATCAATGAGAGACTTTTTCAGCTACTATCGT	1174
DB	901	CAGAGAAATGAAGGTACCTAGAAATCCATCAATGAGAGACTTTTTCAGCTACTATCGT	960
QY	1175	CATTGAAACTGGCCCTTTAAAGGACGAGAGTGAAGGAAT	1216
DB	961	CATTGAAACTGGCCCTTTAAAGGACGAGAGTGAAGGAAT	1002
RESULT 5			
MMKIN17			
LOCUS	Mouse kin17 mRNA for kin17 protein.	1414 bp	mRNA linear ROD 30-OCT-1995
DEFINITION	Mouse kin17 mRNA for kin17 protein.		
ACCESSION	X58472		
VERSION	X58472.1	GI:1045209	
KEYWORDS	kin17 gene; kin17 protein; zinc-finger protein.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;		
	Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 1414)		
AUTHORS	Angulo, J.F., Rouer, E., Benarous, R. and Devoret, R.		
TITLE	Identification of a mouse cDNA fragment whose expressed polypeptide		

JOURNAL PUBMED REFERENCE	reacts with anti-recA antibodies Biochimie 73 (2-3), 251-256 (1991) 1715759	Db	3	AGTTGTAGCGCTCAGGGTCCCATGGGCAAGTCGGATTTTCTGAGCCCCAAGGCCATCGC	62
AUTHORS	Angulo,J.F., Rouer,E., Mazin,A., Mattei,M.G., Tissier,A., Horellou,P., Benarous,R. and Devoret,R.	Qy	104	CAACAGGATCAAGTCCCAAGGGCTGCAGAAGCTACGCTGGTATTGCGAGATGTGCCAGAA	163
TITLE	Identification and expression of the cDNA of KIN17, a zinc-finger gene located on mouse chromosome 2, encoding a new DNA-binding protein	Db	63	CNATAGAAATTAAGTCCAAAGGGCTCCAGAAGCTTCGCTGGTACTGCCAGATGTGCCAAA	122
JOURNAL PUBMED REFERENCE	Nucleic Acids Res. 19 (19), 5117-5123 (1991) 1923796	Qy	164	GCAGTCCGGGACGAGAATGGCTTAAAGTGTCAATGTGTATGTCGAAATCTCATCAGAGACA	223
AUTHORS	Angulo,J.	Db	123	GCAATGCCGCGACGAGAATGGCTTAAAGTGTCACTGTATGTCTGAAATCTCATCAAGACA	182
JOURNAL PUBMED REFERENCE	Direct Submission Submitted (25-MAR-1991) J. Angulo, C N R S, GEMC-Lab d'Enzymologie, 1 Avenue de la Terrasse, Gif-Sur-Yvette, Ile-de-France, FRANCE	Qy	224	ACTATTGCTGGCTTCAGAAAATCCTCAGCAGTATTATGGATATTATTTTCAGAGGAATTCG	283
AUTHORS	revised by [4] MPGP	Db	183	ACTGTGCTGGCTTCAGAAAACCTCAGCAGTATTATGGATATTATTTTCAGAGGAATTCG	242
JOURNAL PUBMED REFERENCE	On Oct 31, 1995 this sequence version replaced gi:52795. Location/Qualifiers	Qy	284	AAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAAAAGGGTCACAACAACAT	343
AUTHORS	Angulo,J.	Db	243	AAATGACTTTCTGGAACTTCTGAGGCGACGCTTTGGCACTAAAAGGGTCCACACACAT	302
JOURNAL PUBMED REFERENCE	Direct Submission Submitted (30-OCT-1995) J. Angulo, C N R S, GEMC-Lab d'Enzymologie, 1 Avenue de la Terrasse, Gif-Sur-Yvette, Ile-de-France, FRANCE	Qy	344	TGTTCTACAACGAATACATCAGCCACCAGAGCACATCCACATGAATGCCATCTCAGTGGGA	403
AUTHORS	On Oct 31, 1995 this sequence version replaced gi:52795.	Db	303	TGTTCTACAATGAATACATCAGCCACCAGAGCACATCCACATGAACGCTACCCAGTGGGA	362
JOURNAL PUBMED REFERENCE	Location/Qualifiers	Qy	404	AACTCTGACTGATTTTACTAAGTGGCTGGGCGAGAGAGGCTTGTCAAAGTGGACGAGAC	463
AUTHORS	1. .1414	Db	363	GACACTGACCGACTTTTACCAAGTGGCTGGGCGAGAGAGGGCTTGTGTAAGTGGATGAGAC	422
JOURNAL PUBMED REFERENCE	/organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /cell_lines="MB66" /tissue_type="brain" /dev_stages="embryo"	Qy	464	ACCAAAAGGCTGGTATATTTCAGTACATAGACAGGAGCCAGAAAATCTATCCCGGCGCACT	523
AUTHORS	1. .1414	Db	423	ACCGAAGGCTGGTACATTCAGTACATAGACAGAGACCCAGAAAACCATCCGTCGCGCACT	482
JOURNAL PUBMED REFERENCE	/gene="KIN17" /pseudo	Qy	524	GGAACTCGAGAAAAAGAAAAAGCAGGACCTTGATGTAGTGAAGAAAAAACTGCAAAATTTAT	583
AUTHORS	<1. .1382	Db	483	GGAAATTAGAAAAAAGAAAGCAAGATCTGGACGATGAAGAAAAAACTGCAAGATTTCAT	542
JOURNAL PUBMED REFERENCE	/gene="KIN17" /pseudo	Qy	584	TGAAGACCAAGTGAGAGAGCCCTGGAAAGGAGGAACAGAGGCTCCCTACTTTTACGGA	643
AUTHORS	25. .1200	Db	543	TGAGGAGCAGGTGAGNAGAGGCCCTGGAAAGGGAAGAGCAGCAGAGACACCTGTTTTACAGA	602
JOURNAL PUBMED REFERENCE	/gene="KIN17" /pseudo	Qy	644	ATTAAGCAGAGAAAAATGATGAAGAGAAAGTCACCGTTTAAATTTGATGAAGAGCAGCATGTAG	703
AUTHORS	/codon_start=1 /product="kin17 protein" /db_xref="PSEUDO:CAA41386.1"	Db	603	ACTTAGCCGAGAAAATCGAGNAGAAAAGTTACGTTCAATCTGAATAAAGGAGCGGGTGG	662
JOURNAL PUBMED REFERENCE	106. .174	Qy	704	CTCATCCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGAGTGCCTGAAGACGAT	763
AUTHORS	/gene="KIN17" /note="zinc-finger of kin17"	Db	663	CTCAGCGGAGCTACAAACATCCAAAGTCAAGCTCTTTGGGACCAAGTGCCTGAAGAGTCT	722
JOURNAL PUBMED REFERENCE	234. .842	Qy	764	AGGAAGTTTCAGCATCAGTGAAGCGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGA	823
AUTHORS	/gene="KIN17" /note="kin17(200) fragment"	Db	723	GGGAGGCGCAGCATCCGGGAAACCGAAAGAGTCTTTCACAGAGCTCCGCCAGCCT-----	777
JOURNAL PUBMED REFERENCE	/evidence=experimental 472. .496	Qy	824	AAAGAAGAAAAAGAAATCTGCACCTGGATGAAATCATGGAGATTTGAAGAGGAAAAAGAAAG	883
AUTHORS	/gene="KIN17" /note="kin17 helix1 motif"	Db	778	-GCCGAGNAGNAGAAAGTCCGCCCTGGATGAGATCATGGAGCTCGAAGAGGAAAGNAGAAAG	836
JOURNAL PUBMED REFERENCE	781. .792	Qy	884	AACTGCCCGAACAGACTACTGGCTACAGCCCTGAAATTTATGTGAAAAATTTATAACCAAGAA	943
AUTHORS	/gene="KIN17" /note="nuclear localization signal 1"	Db	837	GACCGCACGGACAGACGCTGGTTACAGCCGGGGATCGTTGTGAAAAATTTATAACGAAGAA	896
JOURNAL PUBMED REFERENCE	907. .924	Qy	944	ACTGGGAGAGAAATATCATAGAAAGAGGCTATTGTTTAAAGGAAGTAAATGACAAATATAC	1003
AUTHORS	/gene="KIN17" /note="nuclear localization signal 2"	Db	897	GCITGGGGAGAAATATCACAAAGAAAGAGGGGTGTTTAAAGGAAGTATTGACAGGTACAC	956
JOURNAL PUBMED REFERENCE	1382	Qy	1004	AGCTGTTGTGAAGATGATTGATTTCTGGAGACAAGCTGAAACTTGACCAGAGCTCATTTAGA	1063
AUTHORS	/gene="KIN17"	Db	957	AGCTGTGTTAAGATGACTGACTCTGGAGACAGGCTGAAACTGGACAGACTCATTTAGA	1016
JOURNAL PUBMED REFERENCE	65.4%; Score 999.8; DB 9; Length 1414; Best Local Similarity 82.7%; Pred. No. 4e-200; Matches 1158; Conservative 0; Mismatches 237; Indels 6; Gaps 1;	Qy	1064	GACAGTAATTCAGCACACAGGAAAAAGAAATTCATAGTTTAAATCGAGGCTACAGAGGAAA	1123
AUTHORS	44 AGTGATCGCTCCGTGGTGGCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGC	Db	1017	GACAGTCAATTCGCGCCCGGGGAAAAAGGGTTCTAGTTTAAATCGAGGCTACAGAGGAAA	1076
JOURNAL PUBMED REFERENCE		Qy	1124	TGAAGGTACCTTAGAATCCATCAATGAGAGACTTTTTTCAGTACTACTATCGTCATTGAAC	1183
AUTHORS					

[illegible]

Db 1377 AAAAAAAAA 1384

RESULT 7
AX003331

LOCUS AX003331 1390 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 24 from Patent WO929845.
ACCESSION AX003331
VERSION AX003331.1 GI:9927146
KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS Angulo-Mora, J. F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 24 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)

FEATURES
source Location/Qualifiers
1..1390
/organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"

ORIGIN
Query Match 64.8%; Score 990.6; DB 6; Length 1390;
Best Local Similarity 83.0%; Pred. No. 3.5e-198;
Matches 1144; Conservative 0; Mismatches 229; Indels 6; Gaps 1;

QY 66 ATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCACAGGATCAAGTCCAAGGG 125
DB 1 ATGGCAGTCGATTTCTGAGCCCCAAGGCCATCGCCATAGATTAAGTCCAAAGG 60

QY 126 CTCGAGAAGCTACGCTGGTATTGCCAGATGTGCCAAGACGAGTCCCGGACGAGAATGCC 185
DB 61 CTCGAGAAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCCGACGAGAATGCC 120

QY 186 TTTTAAGTGTCATTGTATGTCGGAATCTCATCAGACCACTATTGTCTGGCTTCAGAAAT 245
DB 121 TTTTAAGTGTCATTGTATGTCGGAATCTCATCAAGACAACTGTGTCTGCTTCAGAAAC 180

QY 246 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTCTTAGAACTTCTC 305
DB 181 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTCTTAGAACTTCTG 240

QY 306 AGGAGACGCTTTGGCACTAAAGGGTCCACAAACAATGCTCTACACGAATACATCAGC 365
DB 241 AGGCGACGCTTTGGCACTAAAGGGTCCACAAACAATGCTCTACAAATGAATACATCAGC 300

QY 366 CACCGAGACACATCCACATGAATGCCACTCAGTGGGAACCTCTGACTGATTTTACTAAG 425
DB 301 CACCGAGACACATCCACATGAACGCTTACCAGTGGGAGACCTCTGACCGACTTTTACCAG 360

QY 426 TGGCTGGGCAGAGAAGGCTTTGTCAAAGTGGACGAGACACAAAGGCTCGTATATTTCAG 485
DB 361 TGGCTGGGCAGAGAAGGCTTTGTTAAAGTGGATGAGACACCGAAGGCTCGTATATTTCAG 420

QY 486 TACATGACAGGACCCAGAAAATCTCCGCGGCAACTGGAACTGGAGAAAAAGAAAAAG 545
DB 421 TACATGACAGACACCAGAAAACCATCCGTCGCAACTCGAAATTAGAAAAAGAAAGAAG 480

QY 546 CAGGACCTTGATGATGAAGAAAAAATGCCCCAAATTTATTGAAGACGCAAGTGAAGAGGC 605
DB 481 CAAGATCTGACGATGAAGAAAAAATCTGCCAAGTTCAATTGAGGACGAGGTGAGAAAGGC 540

QY 606 CTCGAGGGGAAGGAACAGGAGGTCCTTACTTTTACGGAAATTAAGCAGAGAAAAATGATGA 665
DB 541 CTCGAGGGGAAGGACGAGGAGACACCTTGTTTTTACAGAACTTAGCCGAGAAATGAGAA 600

QY 666 GAGAAAGTCACGTTTAAATTTTGAAGTAAGGAGCATGTAGCTCATCCGAGCAACATCTTCC 725

Db 601 GAAAGTACCTTCAATCGATTAAGAGCGGTGGCTCAGCGGAGCTACACATCC 660
 QY 726 AAGTCAAGTACTCTGGGACCGAGTGCACTGAAGACGATAGGAAGTTTCAGCATCAGTGA 785
 Db 661 AAGTCAAGCTCTTTGGGACCAAGTGCACTGAAGCTGCTGGGAGGCGCAGCATCCGGGAA 720
 QY 786 CGAAAGATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGCAAGAAAGAAATCTGCA 845
 Db 721 CGAAAGAGTCTTCAAGAGCTTCCGCCAGCTGC-----GAAGAGGAAGATCGGCC 774
 QY 846 CTGGATGAATCATGTGAGATTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 905
 Db 775 CTGGATGAGATCATGTGAGCTCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 834
 QY 906 CTACAGCTGAAATTAATTTGTAAGAAATTAACCAAGAACTGGGAGAGAAATATCATAG 965
 Db 835 TTACAGCCGGGATCGTTGTGAATTAATTAACGAAGAAGCTTTGGGAGAGAAATATCACA 894
 QY 966 AAAAGGCTATTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAAGATGATTGAT 1025
 Db 895 AAGAAAGGGTCTGTTAAGGAAGTAAATTGACAGGTACAGCTGTTGTGAAGATGACTGAC 954
 QY 1026 TCTGGAGACAAGCTGAAACTTGACCAAGCTCATTTAGAGACAGTAAATCCAGCACCAGGA 1085
 Db 955 TCTGGAGACAGCTGAAACTTGACCAAGCTCATTTAGAGACAGTCAATTCGGCCCGGGG 1014
 QY 1086 AAAAGAAATCTAGTTTAAATGAGGCTACAGAGAAATGAAGTACCTAGATCCATC 1145
 Db 1015 AAAAGGGTCTAGTTTAAATGAGGCTACAGAGAAATGAAGGCACTCTCGAATCCATC 1074
 QY 1146 AATGAGAAGCTTTTTCAGCTACTCTGCTGATGAACTGGCCCTTTAAAGAGGACGACGA 1205
 Db 1075 AATGAGAAGCTTTTTCAGCTACTCTGCTGATGAACTGGCCCTTTAAAGAGGACGACGA 1134
 QY 1206 GTTGAAGGAATCAATATGAAGACATTTCTAACTTGCTGAGTTTGAAGAAATTTGTTAAC 1265
 Db 1135 GTTGAAGGATTCATATGAGACATATCTAACTTGCTGAGTTTGAAGAAATTTGTTAAC 1194
 QY 1266 AATACATTAATTTTAAAGATCAAAATGGTGTTCGCAAGGCATATGAGACTTACT 1325
 Db 1195 AACACATTTGAACCTGTGAAGCATCAAAATGGTGTTCGCAAGGCATGTTGTAATCTACT 1254
 QY 1326 GTGTTAGGTATATCTTTTGTATTAACCAAGAGGTTTGTGAATATATCTGTTATAGT 1385
 Db 1255 GTGTTAGGTATATCTTTTGTATTAACCAAGAGGTTTGTGAATATATCTGTTATAGT 1314
 QY 1386 TGTTTCAGCTAACTTTGAGAGAAATTAATATGTTCTCATGAGTATCAAACTATGTAA 1444
 Db 1315 TAGTGGGTAAATTTATATAAATCTATGTTTGTATTAAGTGAAGGAGGAGGAGGAGG 1373

RESULT 8
 AC119716
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-274116, *** SEQUENCING IN PROGRESS
 AC119716
 VERSION AC119716.6 GI:25092366
 KEYWORDS HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Rattus.
 1 (bases 1 to 221547)
 Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Unpublished
 2 (bases 1 to 221547)
 Worley K.C.
 Direct Submission
 Submitted (01-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 221547)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23611286.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWLC
Center clone name: CH230-274116
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 208013 bases at least Q40
Consensus quality: 211685 bases at least Q30
Consensus quality: 213907 bases at least Q20
Estimated insert size: 199182; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 198201: contig of 198201 bp in length
* 198202 198301: gap of unknown length
* 198302 199749: contig of 1448 bp in length
* 199750 199849: gap of unknown length
* 199850 201115: contig of 1266 bp in length
* 201116 201215: gap of unknown length
* 201216 202338: contig of 1123 bp in length
* 202339 202438: gap of unknown length
* 202439 203500: contig of 1062 bp in length
* 203501 203600: gap of unknown length
* 203601 204905: contig of 1305 bp in length
* 204906 205006: gap of unknown length
* 205007 206192: contig of 1187 bp in length
* 206193 206292: gap of unknown length
* 206293 207385: contig of 1093 bp in length
* 207386 207485: gap of unknown length
* 207486 208519: contig of 1034 bp in length
* 208520 208619: gap of unknown length
* 208620 209758: contig of 1139 bp in length
* 209759 209858: gap of unknown length
* 209859 211441: contig of 1583 bp in length
* 211442 211541: gap of unknown length
* 211542 212876: contig of 1335 bp in length
* 212877 212976: gap of unknown length
* 212977 214256: contig of 1280 bp in length
* 214257 214356: gap of unknown length
* 214357 217003: contig of 2647 bp in length
* 217004 217103: gap of unknown length
* 217104 218459: contig of 1356 bp in length
* 218460 218559: gap of unknown length
* 218560 219788: contig of 1229 bp in length
* 219789 219888: gap of unknown length
* 219889 221547: contig of 1659 bp in length.
Location/Qualifiers
1. .221547
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-274116"
1. .1272
/note="wgs_end extension
clone_end:Sp6"
complement(4416..5263)
/note="clone boundary
clone_end:Sp6"
site:
end_sequence:B2266352"
192445..193553
/note="wgs_contig"
195743..196592
/note="clone_boundary

clone_end:T7
site:
end_sequence:B2266349"
198202..198301
/estimated_length=unknown
199750..199849
/estimated_length=unknown
201116..201215
/estimated_length=unknown
202339..202438
/estimated_length=unknown
203501..203600
/estimated_length=unknown
204906..205005
/estimated_length=unknown
206193..206292
/estimated_length=unknown
207386..207485
/estimated_length=unknown
208520..208619
/estimated_length=unknown
209759..209858
/estimated_length=unknown
211442..211541
/estimated_length=unknown
212877..212976
/estimated_length=unknown
214257..214356
/estimated_length=unknown
217004..217103
/estimated_length=unknown
218460..218559
/estimated_length=unknown
219789..219888
/estimated_length=unknown
Query Match 63.4%; Score 969.2; DB 14; Length 221547;
Best Local Similarity 78.6%; Pred. No. 5.6e-194;
Matches 1101; Conservative 0; Mismatches 293; Indels 6; Gaps 1;
QY 49 TCGCTGCCGTGTCGCCATGGGAAGTCGGATTTCTTACTCCCAAGGCTATCCGCAACA 108
DB 212995 TAGCGCTCAGGTCGCCATGGGCAAGTCGGATTTCTTACAGCCCCCAAGCCATCCCAATA 213054
QY 109 GGATCAAGTCCAGGGGCTGCAGAAAGTACGCTGGTATTCAGATGTCAGAGAGAGT 168
DB 213055 GAATTAAGTCCAAAGGGCTCCAGAAAGCTTCGCTGGTACTGCCAGATGTGCCCCAAAGCAAT 213114
QY 169 GCCGGGACGAGAAATGGCTTTAAGTGTCTATTTGATGTCGGAATCTCATCAGAGACAATCTAT 228
DB 213115 GCCGCGACGAGAAATGGCTTTAAGTGTCTATTTGATGTCGGAATCTCATCAGAGACAATCTAT 213174
QY 229 TGCTGGCTTCAGAAAATCCTCAGCAGTTTATGGGATTTATTTTTCAGAGGAATTCGGAATG 288
DB 213175 TGCTGGCTTCAGAAAATCCTCAGCAGTTTATGGGATTTATTTTTCAGAGGAATTCGGAATG 213234
QY 289 ACTTTCTAGAACTTCTCAGGAGACGCTTTGGCAGTAAAGGGTCCACAAACAATTTGTCT 348
DB 213235 ACTTTCTGGAACCTTCTGAGGCGACGCTTTGGCAGTAAAGGGTCCACAAACAATTTGTCT 213294
QY 349 ACAAGGATATCATCAGCCACCGAGAGCACATCCACATGAATGCCACTCAGTGGGAACTC 408
DB 213295 ACAATGAATATCATCAGCCACCGAGAGCACATCCACATGAACGCTACTCAGTGGGAGACAC 213354
QY 409 TGACTGATTTTACTAAGTGGCTGGCGAGAGAAAGGCTTTGTGCAAAAGTGGACGAGACACCA 468
DB 213355 TGACCGACTTTTACCAAGTGGCTGGCGAGAGAGGGCTTTGTGTAAAGTGGATGAGACACCGA 213414
QY 469 AAGCTGGTATATTTCAGTACATAGACAGGGAGCCAGAAAATCTATCCGCCGGCAACTGGAAAC 528
DB 213415 AAGCTGGTATATTTCAGTACATAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 213474
QY 529 TGGAGAAAAGAAAAGACGAGACCTTGTATGATGAAGAAAAAACTGCCCAATTTTATTGAAG 588

Db	213475	TGAAAAAAGAGAGACGAGATCTGGACGATGAGAAAAAAGCTCCAGTTCAATGAGG	213534
QY	589	AGCAAGTGAGAGAGCGCTTGGAGGAGGAAAGGAAACAGAGGTCCTACTTTTACGGAATAA	648
Db	213535	AGCAGTGAGAGAGCGCTTGGAGGAGGAAAGAGAGCAGAGACACCTGTTTACAGAACTAA	213594
QY	649	GCAGAGAAATGATGAGAGAAAGTCACGTTTAAATTTGAGTAAAGGACATGACTCAT	708
Db	213595	GCCGAGAAATGAGGAGAGAAAAAGTTACGTTCAATCTGAATAAAGAGCGGGTGCCTCAG	213654
QY	709	CCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGACCGAGTGCACCTGAAAGACATAGAA	768
Db	213655	CCGGAGCTACAACATCCAAAGTCAAGTCTTTGGACCAAGTGCATGAAAGTCTGCGGGA	213714
QY	769	GTTCAGCATCAGTGAAGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGA	828
Db	213715	GCGCAGATCCGGGAAACGAAAGAGTCTTACAGAGCTCCGCCAGCCTGC-----GA	213768
QY	829	AGAAAAAGAAATCTGCACCTGGATGAATCATCGAGATTGCAAGGAGGAAAGAAAGAACTG	888
Db	213769	AGAAGAAGAAGTCCGCCCTGGATGATCATGGAGCTCGAAGGAGGAAAGAAAGGACCG	213828
QY	889	CCCGACACACTACTGCTACAGCTGAAATTTATTTGTGAAATTTATTAACCAAGAACTCG	948
Db	213829	CACGACAGACGCTGTTACAGCCGGGATCGTTGTGAAATTTATTAACCAAGAACTCG	213888
QY	949	GAGAGAAATATCATAGAAAAAGGCTATTTGTAAGGAAGTAAATGTAACAATATACAGCTG	1008
Db	213889	GGGAGAAATATCAAGAAAGAAAGGGTCTGTTAAGGAAGTGTGACAGGTACACAGCTG	213948
QY	1009	TTGTGAAGATGATGATTTCTGGAGACAAGCTGAAACTTGACCAAGTCAATTTAGAGACAG	1068
Db	213949	TGTTAAGATGACTGACTCTGGAGACAGGCTGAAACTGGAACCAAGTCAATTTAGAGACAG	214008
QY	1069	TAATTCAGCAGCAGGAAAAAGAAATCTAGTTTAAATCGAGGCTACAGAGAAATGAAG	1128
Db	214009	TCATTCGGCCCCGGGAAAGGGTCTAGTTTAAATGGAGGCTACAGAGGAAATGAAG	214068
QY	1129	GTACCCCTAGAAATCCATCAATGAGAAGCTTTTTCAGTACTACTATGCTATGAACTGGCC	1188
Db	214069	GCACTCTCGAATCCATCAATGAGAAGGCTTTTTCAGCCAGATGATCAATGAACTGGAC	214128
QY	1189	CTTTAAAGGAGCCAGAGTTGAAGGAATTCATATGAGACATTTCTAACTTGCCTGAG	1248
Db	214129	CTTTGAAAGGAGCCAGAGTTGAAGGATTTCAATATGAGACATATCTAACTTGCCTGAG	214188
QY	1249	TTTGAATTTGTTAAACATACATTAATAATCTTAAAGCATCAATTTGGTGTTCGCCAAGG	1308
Db	214189	TTTGAATTTGTTAAACATACATTAATAATCTTAAAGCATCAATTTGATGTTAGCCAAGG	214248
QY	1309	CATTATGAGACTCTACTGTGTTAGGGTATATCTTTTGTATAAAACAAACAGGTTTTTGA	1368
Db	214249	CACGTGTNN	214308
QY	1369	AAATATTAATGATGATGTTGTTACGCTTAACTTTGAGAAGAAATTAATTAATGATGATGAG	1428
Db	214309	NN	214368
QY	1429	GTATCAAACTATGTAATTTT 1448	
Db	214369	ATATAACAGGAGCAATATT 214388	

RESULT 9
LOCUS CQ722054 1117 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 7988 from Patent WO02068579.
ACCESSION CQ722054
VERSION CQ722054.1 GI:42282911
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homindae; Homo.

```

QY 807 TCNACTCAGCTTAAGAAAAGAGAAAAGAAAGAAATCTGCACTGGATGAATCATGAGATT 866
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 740 TCAACTCAGCTTAAAGAAAAGAGAAAAGAAAGAAATCTGCACTGGATGAATCATGAGATT 799
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 867 GAAGAGGAAAAGAAAAGAAATGCCCCGAGACAGACTACTGGCTACAGCTTGAATATTGTTG 926
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 800 GAAGAGGAAAAGAAAAGAAATGCCCCGAGACAGACTACTGGCTACAGCTTGAATATTGTTG 859
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 927 AAAATTATAACAAGAAATCGGAGAGAAATATATAGAAAAGGCTATTGTTAAAGGAA 986
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 987 GTAATTGACAAATATACAGCTGTTCTGAAGATGATTCTGGAGACAAGCTGAACTT 1046
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 920 GTAATTGACAAATATACAGCTGTTCTGAAGATGATTCTGGAGACAAGCTGAACTT 979
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1047 GACCAGACTCATTATAGACACAGTAATTCAGCACCAGGAAAAGAAATCTAGTTTAAAT 1106
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 980 GACCAGACTCATTATAGACACAGTAATTCAGCACCAGGAAAAGAAATCTAGTTTAAAT 1039
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1107 GGAGCTACAGAGGAAATGAAGGTACCTAGATCAATCAATGAGAAAGCTTTTCAGCT 1166
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1040 GGAGCTACAGAGGAAATGAAGGTACCTAGATCAATCAATGAGAAAGCTTTTCAGCT 1099
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1167 ACTATCGTCATTGAACT 1184
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1100 ACTATCGTCATTGAACT 1117
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 10
AC166154 130262 bp DNA linear HTG 26-JUL-2005
LOCUS Mus musculus chromosome 8 clone RP24-344F2 map 8, WORKING DRAFT
DEFINITION AC166154
ACCESSION AC166154
VERSION AC166154.1 GI:71143327
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 130262)
Biren,B., Nusbaum,C. and Lander,E.
Mus musculus chromosome 8, clone RP24-344F2
Unpublished

2 (bases 1 to 130262)
Biren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J.,
Choepe,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Fato,S., Ferreira,P., FitzGerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hegopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,J., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Rhee,C., Rogov,P., Roman,J., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-JUL-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

```

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Broad Institute of MIT and Harvard
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence submissions@broad.mit.edu
----- Project Information
Center project name: L33627
Center clone name: 344_F_2
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; version 0.960731
Assembly program: Phrap; version 0.960731
Consensus quality: 128346 bases at least Q40
Consensus quality: 129716 bases at least Q30
Consensus quality: 129103 bases at least Q20
Insert size: 20000; agarose-fp
Quality coverage: 10.2 in Q20 bases; agarose-fp
Quality coverage: 15.7 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 39772: contig of 39772 bp in length
* 39773: gap of unknown length
* 39872: contig of 1692 bp in length
* 41564: gap of unknown length
* 41664: contig of 6156 bp in length
* 47820: gap of unknown length
* 47920: contig of 32957 bp in length
* 80977: gap of unknown length
* 80978: contig of 25585 bp in length
* 106562: gap of unknown length
* 106663: contig of 23600 bp in length.
FEATURES
Location/Qualifiers
1..130262
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="8"
/map="8"
/clones="RP24-344F2"
/clone_lib="RPCI-24 Male Mouse BAC"
1..39772
/notes="assembly_fragment"
/clone_end:SP6
/vector_side:left
39773..39872
/estimated_length=unknown
39873..41564
/notes="assembly_fragment"
41565..41664
/estimated_length=unknown
41665..47820
/notes="assembly_fragment"
47821..47920
/estimated_length=unknown
47921..80877
/notes="assembly_fragment"
80878..80977
/estimated_length=unknown
80978..106562
/notes="assembly_fragment"
106563..106662
/estimated_length=unknown
106663..130262
/notes="assembly_fragment"
misc_feature
misc_feature
gap
misc_feature
gap
misc_feature
gap
misc_feature
gap
misc_feature
gap
misc_feature
gap
misc_feature

```

ORIGIN	clone_end:T7 vector_side:right	Query Match Best Local Similarity Matches 1060; Conservative	49.7%; 77.1%; 0;	Score 759; Pred. No. 1.2e-149; Mismatches 260;	DB 14; Indels 55;	Length 130262; Gaps 9;
Qy	46	TGATCGCTCGCTGGTGGCCATGGGGAAGTCGGATTTTCTTACTCCCAAGGGTATCGGCA	105			
Db	99745	TGGTCGGGGCCAGGGTAGCCTTGTGCAAGCGG-TTTTCTTAGCCCTAAGGCGCATCGCA	99803			
Qy	106	ACAGGA-TCAAGTCCAAAGGGGTCGAGAGCTACGCTGGTATTCGACAGATGGCCAGAG	164			
Db	99804	ATAGAAATGAAGTCCAAAGGGGTCGAGAGCTTCAACGGTACTGCCAGATGGCCAAAG	99863			
Qy	165	CAGTGGCGGACAGAGATGCTTTAAGTGTCTATGTCGGAATTCATCAGAGACAA	224			
Db	99864	CAATGCCGCCACAGAGATGTCTTTAAGTGTCACTGTATGTCTTAATCTCATCAAGACCA	99923			
Qy	225	CTATTGCTGCTTCAGAAATCCTCAGCAGTTTATGGATTTATTTTTCAGAGGAATTCGGA	284			
Db	99924	CAGTTGCTGCTTCAGAAAGCCTCAGCAGTTTATACATTTATTTTCAG-----	99973			
Qy	285	AATGACTTTCAGAACTTCTCAGGAGACGCTTTGGCCTAAAGGGTCCAGAACACATT	344			
Db	99974	-----GCATTAAGGGGCCCAACAACATT	99999			
Qy	345	GTCTACAAAGATACATCAGCCACGAGAGCATCCACATGAATGCCACTCAGTGGGAA	404			
Db	100000	GTCTACAAATGAATATATCAGCCACCAAGAGCATCCACATGAATGCTACCCAGTGGGAA	100059			
Qy	405	ACTCTGACTCATTTTACTAAGTGGCTGGGCGAGAGAGGCTTTGTCAAAGTGGACGAGACA	464			
Db	100060	AAGTGCACCACTTTCCAGATGGCTGGGCGAG-AGGGCTTGTGTAAGTGGGTGAGACA	100118			
Qy	465	CCAAAAGGCTGGTATATTCATAGTACATAGACAGGAGCCAGAACTATCCCGCGCACTG	524			
Db	100119	CCAAAAGGCTGATACA-TCTGTACATAGACAGAGACCCAGAACTATCTGTGGGCACTG	100177			
Qy	525	GAACT-CGAGAAAGAAAGCAGACCTTGTATGATGAAGAAACCTGCCAAATTTAT	583			
Db	100178	AAATTAGAAAGAAAGAAAGATGGAAGATCTTTGACGATGAAGAAACCTGCCAAATTCAT	100237			
Qy	584	TGAAGACCAAGTGAAGAGAGGCTTGAAGGAAAGGAAACAGGAGGTCCT-ACTTTTACGG	642			
Db	100238	TGAGAGCAGGTGAGAGAGGCTTGGAGGAAAGAGCAGGAGACCTGTTTTTACAG	100297			
Qy	643	AATTAGCAGAGAAATATGATGAAGAGAACTACGTTTTAATTGATGAAGAGCATGTA	702			
Db	100298	AATTAGCAGAGAAATAAGAGAGAAATTTATGTCATCTGAATGAAGAGGAGTGGTG	100357			
Qy	703	GCTCATCGGAGCAACATCTTCCAGTCAAGTACTCTGGGACCGAGTGCAGTCAAG	762			
Db	100358	GCTCAGCAGGAGCTACAAAGTCCAGGTCACAGTCTTTTGGACCAAGTGCAGTCAAGCTGC	100417			
Qy	763	TAGGAAGTTCAGCATCAGTGAAGCAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAG	822			
Db	100418	TGAGAGCAGACATCCGGGAAATGGAAAGGCTTTCACAGAGCTCTCCAGCTCT--G	100474			
Qy	823	AAAAGAGAAAGAAATCTGCATCGATGAAATCATGGAGATGAAGAGGAAAGAA	882			
Db	100475	CAAGAGAGAGAGATGCTGCTGCTGATGAATTTATAGAGATTAAGAGAGAAAGAA	100534			
Qy	883	GAACTGCCGAAACAGACTCTGGCTCAGCCTGAAATTTTGTGAAATTTATTAACCAAG	942			
Db	100535	GGACCACACGACAGACTCTCTGGTTACAGCCGGGAATCATTTGTGAAACGATTAAGGAAG	100594			
Qy	943	AACCTGGGAGAGAAATATCATAGAAAGAGGCTATTTGTTAAGGAGATTAATGACAAATATA	1002			
Db	100595	AGCTTGGGAGAGATATCATAGAAAGAAAGGGGTCATTACGGAAGAGATTCACATGTACA	100654			
Qy	1003	CAGCTGTTGTGAAGATGATTTCTTGAGACAAGCTGAAACTTTGACAGACTCATTTAG	1062			

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (<http://www.chori.org>) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES

source	1. 140215	
	/organism="Mus musculus"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:10090"	
	/chromosome="8"	
	/map="8"	
	/clone="RP24-344P2"	
	/clone_lib="RPCI-24"	
repeat_region	32..81	/rpt_family="ERVK"
repeat_region	333..426	/rpt_family="ERVK"
repeat_region	432..600	/rpt_family="ERVK"
repeat_region	605..763	/rpt_family="ERVK"
repeat_region	830..1157	/rpt_family="ERVK"
repeat_region	1158..1387	/rpt_family="ERVK"
repeat_region	1426..1641	/rpt_family="ERVK"
repeat_region	1657..1779	/rpt_family="ERVK"
repeat_region	2071..2147	/rpt_family="ERVK"
repeat_region	2242..2391	/rpt_family="ERVK"
repeat_region	2832..2881	/rpt_family="ERVK"
repeat_region	3133..3226	/rpt_family="ERVK"
repeat_region	3232..3400	/rpt_family="ERVK"
repeat_region	3405..3563	/rpt_family="ERVK"
repeat_region	3630..3957	/rpt_family="ERVK"
repeat_region	3958..4187	/rpt_family="ERVK"
repeat_region	4226..4441	/rpt_family="ERVK"
repeat_region	4457..4579	/rpt_family="ERVK"
repeat_region	4871..4947	/rpt_family="ERVK"
repeat_region	5042..5191	/rpt_family="ERVK"
repeat_region	5632..5681	/rpt_family="ERVK"
repeat_region	5933..6026	/rpt_family="ERVK"
repeat_region	6032..6200	/rpt_family="ERVK"
repeat_region	6205..6363	/rpt_family="ERVK"
repeat_region	6430..6757	/rpt_family="ERVK"
repeat_region	8931..9312	/rpt_family="L1"
repeat_region	9434..9660	/rpt_family="L1"
repeat_region	9682..9931	/rpt_family="L1"
repeat_region	9959..10098	/rpt_family="L1"
repeat_region	10770..10883	/rpt_family="L1"
repeat_region	11090..11220	/rpt_family="Alu"
repeat_region	11240..11299	/rpt_family="L1"
repeat_region	11704..11890	/rpt_family="ERV1"
repeat_region	11977..12288	/rpt_family="ERV1"
repeat_region	13460..13519	/rpt_family="ERV1"
repeat_region	13586..13801	/rpt_family="RMR10B"
repeat_region	18269..18413	/rpt_family="RMR10B"
repeat_region	19731..21946	/rpt_family="ERV1"
repeat_region	21953..23462	/rpt_family="ERV1"
repeat_region	23875..24058	/rpt_family="ERV1"
repeat_region	24558..24711	/rpt_family="ERVK"
repeat_region	25065..25121	/rpt_family="L1"
repeat_region	26157..26279	/rpt_family="RMR10A"
repeat_region	27231..27425	/rpt_family="RMR10A"
repeat_region	27505..27718	/rpt_family="RMR10A"
repeat_region	27956..28084	/rpt_family="MaLR"
repeat_region	28245..28470	/rpt_family="MaLR"
repeat_region	30676..30798	/rpt_family="MaLR"
repeat_region	31586..31640	/rpt_family="L1"
repeat_region	31634..32651	/rpt_family="L1"
repeat_region	33734..33932	/rpt_family="MER1_type"
repeat_region	35881..36272	/rpt_family="MaLR"
repeat_region	38599..39351	/rpt_family="L1"
repeat_region	40422..40554	/rpt_family="L1"
repeat_region	41164..41266	/rpt_family="B2"
repeat_region	42885..43336	/rpt_family="L1"
repeat_region	45098..45157	/rpt_family="ERV1"
repeat_region	45351..45480	

Db	82524	GGACCACACGGACAGACTCTCTGTTTACAGCGGGGAATCATTTGTGAACGTTATAGGAAGA	82583
Qy	943	AATCGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAGTANTTGCAGAAATATA	1002
Db	82584	AGCTTGGGAGAGATATCAACAAGAAAAAAGGGGTCATTTACGAGAGAGATTGACATGTACA	82643
Qy	1003	CAGCTGTTGTGAAGATGATTGATCTTGAGAGACAAGCTGAAACTTGACACAGACTCATTTTAG	1062
Db	82644	CAGCTGTTGTGAAGTGTGACTCTTGAGACACAGGCTGAACTGACACAGACTCATTTTAG	82703
Qy	1063	AGACAGTAATTCACGACACAGGAAAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAGAA	1122
Db	82704	AGACAGTCATTCGGGCCCGAGGCAAGGGTCTAGTTTTTAAATGGAGGCTACAGAGAA	82763
Qy	1123	ATGAAGGTACCTTAGAATCCATCAATGAGAAAGCTTTTTTCAGCTACTATCGTCATTGAAA	1182
Db	82764	ATGAAGGCACCTTCAAATACATCAATGAGAAAGGCTTTTTTCGGCTAAGATAGTCATTGAAA	82823
Qy	1183	CTGGCCCTTTAAAGAGACGACAGTTTGAGGAAATTCATATATGAGACACATTTCTTAAACTTG	1242
Db	82824	TTGACCTTTGAAAGAACACAGATAGAACGTATTCATATATGAGACATATCTATCTTG	82883
Qy	1243	CCTGAGTTTGAAATTTGTTTAAACAATACATTTAAATCTTTAAAGCATCAAAATGGTTCG	1302
Db	82884	CTTGAGTTTGAAATTTGTTAGAACACATTCGAAATTTGAAAGCATCAAAATGGTTCG	82943
Qy	1303	CCAAGGCATATGAGACTCTACTGTTGTTAGGTATATCTTTTGTATATAAACAACAGGT	1362
Db	82944	TCAAGGCATTTGTAACCTCTACTGTTGTTAGGGAATTTGTTTGTATATAAAAAATCATCT	83003
Qy	1363	TTTTGAAATATTTACTGTATAGTTGTTTCAGCTAAACTTTTGAGAGAAATTTTAATTA	1417
Db	83004	ATTT--AAGTAATCTAAATAATTCGTTAAATAAATCTTATGTTATTTTTTTAA	83056
RESULT 12			
LOCUS	BC074606	1483 bp	linear
DEFINITION	Xenopus tropicalis MGC69184 protein, mRNA	(cdna clone MGC:69184	
	IMAGE:53071170), complete cds.		
ACCESSION	BC074606		
VERSION	BC074606.1	GI:49257753	
KEYWORDS	MGC.		
SOURCE	Xenopus tropicalis (Silurana tropicalis)		
ORGANISM	Xenopus tropicalis		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Silurana.		
AUTHORS	1 (bases 1 to 1483)		
	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D., Dickinson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalek, U., Smalls, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
PUBMED	proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
REFERENCE	12477932		
	2 (bases 1 to 1483)		

AUTHORS Klein, S. and Gerhard, D.S.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-2004) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA
REMARK NIH-MGC Project
COMMENT Contact: XGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Robert M. Grainger
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: DOE Joint Genome Institute, Walnut Creek, California
 Web site: www.jgi.doe.gov

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAX Plate: 130 Row: b Column: 23
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, similarity but not identity to protein.

FEATURES

Location/Qualifiers
 1..1483
 /organism="Xenopus tropicalis"
 /mol_type="mRNA"
 /db_xref="taxon:8364"
 /clone="MGC:69184 IMAGE:5307170"
 /tissue_type="Embryo, Silurana tropicalis, gastrula (stages 10-13)"
 /clone_lib="NICHG_XGC_Emb5"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6.1"
 1..1483
 /gene="MGC69184"
 /db_xref="GeneID:448079"
 59..1228
 /gene="MGC69184"
 /codon_start=1
 /product="MGC69184 protein"
 /protein_id="AAH74606.1"
 /db_xref="GI:49257754"
 /db_xref="GeneID:448079"
 /translation="MGRADFLTPKAIANRISKLGKLRWYQMCQKQCRDENGFKCH
 CMSHSORQLLASNPQDFYSDERTEFLRLRRFQTKRVHNNI VVNEYISHR
 EHVHNAQTQWETLDFTKWLGRAGYCKVDETPKGYI QYIDRDEPTIRKQEQEKKK
 QDLDEERTAKFIIEQVKGLLEVQNTPTVTELSRQNEEKVAFNLNKGASTSGTAT
 TKSSSLVSNALKSAVMGSVKR KDTSOVQPKKKKSGALDEIMEMEKKKKTERTDFW
 LQPEIVKIVTKLGEKKYKKKAVVKEVIDRYTAVVKLVDSGDKLKDOTHLETVIPA
 PKGRFVLNGYKGFHGLDCINKEKSFATIEIDSGPLKGRKVGIPYEDISKVA"

ORIGIN

Query Match 46.1%; Score 705; DB 5; Length 1483;
 Best Local Similarity 75.4%; Pred. No. 5,3e-138;
 Matches 908; Conservative 0; Mismatches 285; Indels 12; Gaps 2;
 Qy 54 GCCGTGTCGCATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCCCAACAGGATC 113
 Db 47 GCAGGCTTGGCCATGGGGAAGCAGACTTCCTTACTCCGAGGCAATCGCAATCGTATC 106
 Qy 114 AAGTCCAGGGGCTGCAGAACTAGCTGGTATTCGCGAATGTCGCGAAGCAGTGC 173
 Db 107 AAATCGAAAGGACTCCAGAACTCGGATGGTACTGTCAAATGTGTGAGAAACAGTGT 166
 Qy 174 GACGAGAATGGCTTTAAGTGTCAATGTATGTCGGAATCTCATCAGAGCAACTATTGCTG 233
 Db 167 GATGAGAATGGCTTTAATGCCACTGCATGCTGAGTCGACCAAGCAACTGCTCCTT 226
 Qy 234 GCTTCAGAAAATCCTCAGCAGTTTATGGAATATTTTTCAGAGGAATTCGGAATGACTTT 293
 Db 227 GCTTCTGAAAATCCTCAGCAGTTTATGGAATATTTTTCAGATGAATTTTCGACAGAA 286
 Qy 294 CTAGAACTTCTCAGGAGAGCGCTTTGGCACTAAAGGCTCCACACAAACATTTGTCTACA 353

Db 287 TTGAATTTGCTAAGCGCTCGGTTTGGAAACCAAGAGATACATATAATATTGTCTACAAT 346
 Qy 354 GAATACATCAGCCACCCAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACT 413
 Db 347 GAATATATCAGCCATAGGAGCATGTACACATGAATGCCACCCAGTGGGAAACACTTACA 406
 Qy 414 GATTTTACTAAGTGGCTGGGCGAGAGAGGCTTTGTGCAAGTGGACGAGACACCAAAAGGC 473
 Db 407 GACTTCCCAAAATGGCTTGGCCGAGCAGGTACTGTAAAGTGGATGAGACTCCAAAGGGC 466
 Qy 474 TGGTATATTTCAGTACATAGACAGCGGCCACAGAACTATCCGCCGCAACTTGGAACTGGAG 533
 Db 467 TGGTACATTTCAGTATATAGACAGAGATCCAGAAACAATTCGMAACAGCAGGAACAGAA 526
 Qy 534 AAAAAGAAAAGCAGGACCTTGTATGATGAAGAAAAAACTGCCAAATTTATTGAAGAGCAA 593
 Db 527 AAGAAGAAAACAAGATCTAGATGATGAAGAGAGACTGCTAAATTTATTGAGGAACAA 586
 Qy 594 GTGAGAAGAGCGCTGGAAGGGAAGAAACAGAGAGTCCCTACTTTTACGGAAATTAAGCAGA 653
 Db 587 GTTAAAAAGGGCTTGGAAAGTCACTGAACAGAACACCCCTACCTACACTGAATTTAAGCAGA 646
 Qy 654 GAAATATCATGAAGAGAAAGTCAGCTTTTAAATTTTGAGTAAAGGAGCATCTAGCTCATCCGA 713
 Db 647 CAGAAATGAGGAAGAGAAAGTTGCATTCATTTTAAACAAAGAGCA---AGCACTTCAGGA 703
 Qy 714 GCAACATCTTCCAGTCAAGTACTCTGGGACCGAGTGCACCTGAAGACGATAGGAAGTTCA 773
 Db 704 ACAGCTACAACTAAGTCAAGTCTCTTTGTTATCCAATGCATTAAGTCTGCTGCTTTATG 763
 Qy 774 GCATCAGTGAACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAGAAAGAA 833
 Db 764 GGCTCAGTAAACGAAAGACACCTCTCAGGTGCAACC-----AAAGGAAAAGAAAG 814
 Qy 834 AAGAAATCTGCACCTGGATGAAATCATCGAGATGAAGAGGAAAAAGAAAGAACTGCCCGA 893
 Db 815 AAGAAATCTGCCCTTGTATGAGATTATGGAGATGAAGAGAAAAAGAAAGAGACAGAGCGC 874
 Qy 894 ACAGACTACTGGCTACAGCCTGAAATTTATTGTGAAAAATTATAACCAAGAACTCGGGAGAG 953
 Db 875 ACAGATTTTGGTTTACAGCCTGAAATTTGTTGTAATAAATTTGTACGAAAAAATTTGGGAGAA 934
 Qy 954 AAATATCATTAAGAAAAAGGCTATTGTTAAGGAAGTAAATTCACAATATACAGCTGTTGTG 1013
 Db 935 AAATACTACAAGAAAAAAGCTGTAGTAAAGGAAGTGAATGACAGATACACTGCACTGTTGTG 994
 Qy 1014 AAGATGATTGATTTCTGGAGACAAGCTGAAACTTTGACCAGACTCAATTTAGAGACAGTAAT 1073
 Db 995 AAATCTGTTGATTTCTGGAGATAAATTAAGCTGGATCAGACCCCATTTAGAGACAGTTATC 1054
 Qy 1074 CCAGCACCAGAAAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAGAAATTAAGAGTTACC 1133
 Db 1055 CCAGCACCAGAAAAAGAGAGTTTTTTGTTTTTAAATGGTGTGTTTACAAAGGACATTTTGGCATC 1114
 Qy 1134 CTAGAATCCATCAATGAGAAAGACTTTTTTCACTACTATCTGCTAATTGAACTGGCCCTTTA 1193
 Db 1115 CTAGATTGTCATCAATGAGAAAGAGCTTTTCGGCTACAAATAGAAAATTTGACTCAGAGCACTTA 1174
 Qy 1194 AAAGGACGAGAGTTGAAGGAATTTCAATATGAAGACATTTTCTAACTTGTCTGATTTGA 1253
 Db 1175 AAAGGACGGAAGTGGAAAGCATCCCATATGAAGACATATCCAAAGTTGCTTAGATGTAT 1234
 Qy 1254 AAATT 1258
 Db 1235 TAAAT 1239

RESULT 13
 AC141995/c
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-293H7, *** SEQUENCING IN PROGRESS
 AC141995 144118 bp DNA linear HTG 21-MAR-2003
 ***, 48 unordered pieces.

AC141995	Summary Statistics
AC141995.1 GI:29135466	Sequencing vector: Plasmid;
HTG; HTGS PHASE1	Chemistry: Dye-terminator Big Dye; 100% of reads
Rattus norvegicus (Norway rat)	Assembly program: Phrap; version 0.990329
Rattus norvegicus	Consensus quality: 108995 bases at least Q40
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	Consensus quality: 114645 bases at least Q30
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;	Consensus quality: 119503 bases at least Q20
Sciurionathi; Muroidae; Muridae; Murinae; Rattus	Estimated insert size: 114401; sum-of-contigs estimation
1 (bases 1 to 144118)	Quality coverage: 2x in Q20 bases; sum-of-contigs estimation
Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,	* NOTE: Estimated insert size may differ from sequence length
Allan, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D.,	(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,	* NOTE: This is a 'working draft' sequence. It currently
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,	* consists of 48 contigs. The true order of the pieces
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,	* is not known and their order in this sequence record is
Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E.,	* arbitrary. Gaps between the contigs are represented as
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,	* runs of N, but the exact sizes of the gaps are unknown.
Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,	* This record will be updated with the finished sequence
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,	* as soon as it is available and the accession number will
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,	* be preserved.
Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,	
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,	1 1394: contig of 1394 bp in length
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,	1395: gap of unknown length
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,	1495: contig of 1356 bp in length
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,	2851 2950: gap of unknown length
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,	2951 3972: contig of 1022 bp in length
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,	3973 4072: gap of unknown length
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,	4073 5309: contig of 1237 bp in length
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,	5310 5409: gap of unknown length
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebirt, D., Jackson, A.,	5410 5564: contig of 1155 bp in length
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,	5565 6664: gap of unknown length
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,	6665 8450: contig of 1785 bp in length
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,	8450 8550: gap of unknown length
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,	8550 9904: contig of 1354 bp in length
Lorenshewa, L., Loulsegged, H., Lozado, R.J., Lu, X., Ma, J.,	9904 10004: contig of 1743 bp in length
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,	10004 11847: gap of unknown length
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,	11847 13519: contig of 1672 bp in length
Mawhiney, S., McLeod, M., McNeil, T., Meenen, E., Milosavljevic, A.,	13519 13618: gap of unknown length
Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,	13618 14970: contig of 1352 bp in length
Morris, S., Munidasa, M., Murphy, M., Nair, J., Nankervis, C., Neal, D.,	14970 15070: gap of unknown length
Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G.,	15071 16373: contig of 1303 bp in length
Olarunsgaon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,	16374 16473: gap of unknown length
Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A.,	16474 18359: contig of 1886 bp in length
Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E.,	18360 18460: gap of unknown length
Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,	18460 19822: contig of 1362 bp in length
Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,	19822 19922: gap of unknown length
Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S.,	19922 21450: contig of 1528 bp in length
Scott, G., Shateman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,	21450 23269: contig of 1720 bp in length
Sisson, I., Sitter, C.D., Smales, D., Sneed, A., Sodergren, E.,	23270 23370: gap of unknown length
Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,	23370 25711: contig of 2341 bp in length
Svatek, A., Tabori, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,	25711 25810: gap of unknown length
Tingey, A., Trejos, Z., Umami, K., Valas, R., Vera, V., Villaseana, D.,	25811 27760: contig of 1950 bp in length
Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,	27761 29815: contig of 1955 bp in length
Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R.,	29816 31213: gap of unknown length
Wooden, H., Woley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,	31213 31213: contig of 1298 bp in length
Yen, J., Yoon, H., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,	31214 33042: contig of 1729 bp in length
Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R.,	33043 33143: gap of unknown length
Holt, R.A., Smith, H.O., Weinstein, G. and Gibbs, R.A.	33143 34194: contig of 1052 bp in length
Direct Submission	34195 34294: gap of unknown length
Unpublished	34295 36020: contig of 1726 bp in length
2 (bases 1 to 144118)	36021 37671: contig of 1551 bp in length
Worley, K.C.	37672 37772: gap of unknown length
Direct Submission	37772 39759: contig of 1988 bp in length
Submitted (21-MAR-2003) Human Genome Sequencing Center, Department	39760 42442: contig of 2583 bp in length
of Molecular and Human Genetics, Baylor College of Medicine, One	42443 44482: contig of 1940 bp in length
Baylor Plaza, Houston, TX 77030, USA	44483 44583: gap of unknown length
----- Genome Center	44584 46351: contig of 1769 bp in length
Center: Baylor College of Medicine	
Center code: BCM	
Web site: http://www.hgsc.bcm.tmc.edu/	
Contact: hgsc-help@bcm.tmc.edu	
----- project information	
Center project name: GRGN	
Center clone name: CH230-293H7	

```

* 46352 46451: gap of unknown length
* 46452 48712: contig of 2261 bp in length
* 48713 48812: gap of unknown length
* 48813 51252: contig of 2440 bp in length
* 51253 51352: gap of unknown length
* 51353 54232: contig of 2880 bp in length
* 54233 54332: gap of unknown length
* 54333 57136: contig of 2804 bp in length
* 57137 57236: gap of unknown length
* 57237 59833: contig of 2597 bp in length
* 59834 62619: gap of unknown length
* 62620 62719: gap of unknown length
* 62720 65087: contig of 2368 bp in length
* 65088 65187: gap of unknown length
* 65188 68329: contig of 3142 bp in length
* 68330 68429: gap of unknown length
* 68430 72825: contig of 4396 bp in length
* 72826 72925: gap of unknown length
* 72926 76196: contig of 3271 bp in length
* 76197 76296: gap of unknown length
* 76297 80327: contig of 4031 bp in length
* 80328 80427: gap of unknown length
* 80428 83596: contig of 3169 bp in length
* 83597 83696: gap of unknown length
* 83697 88053: contig of 4357 bp in length
* 88054 88153: gap of unknown length
* 88154 91623: contig of 3470 bp in length
* 91624 91723: gap of unknown length
* 91724 96501: contig of 4778 bp in length
* 96502 96601: gap of unknown length
* 96602 101193: contig of 4592 bp in length
* 101194 101293: gap of unknown length
* 101294 106768: contig of 5475 bp in length
* 106769 106868: gap of unknown length
* 106869 111636: contig of 4768 bp in length
* 111637 111736: gap of unknown length
* 111737 117842: contig of 6106 bp in length
* 117843 117942: gap of unknown length
* 117943 130224: contig of 12282 bp in length
* 130225 130324: gap of unknown length
* 130325 144118: contig of 13794 bp in length.
FEATURES
    Location/Qualifiers
        1..144118
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-293H7"
        1395..1494
            /estimated_length=unknown
        2851..2950
            gap
            gap

Query Match      44.8%; Score 685; DB 14; Length 144118;
Best Local Similarity 80.7%; Pred No. 4,6e-134;
Matches 827; Conservative 0; Mismatches 190; Indels 8; Gaps 2;

Qy 439 AAGCCTTGCAAAAGTGACGACACCAAAAGGCTGGTATATTACGTACATAGACAGG 498
Db 121995 AAGGCTTGTAAGTGTAGACACACCAAAAGGCTGGTATATTACGTACATAGACAGAG 121936
Qy 499 ACCCAGAACTATCCCGCGCACTGGAACTGGAGAAAAAGAACGAGCCTTGATG 558
Db 121935 ACCCAGAACTATACGTCCGCACTGGAACTTAGAAAAAAGAAAGCAAGATCTTGACG 121876
Qy 559 ATGAAGAAAAAAGTCCCAAAATTTATTGAAGAGCAAGTGAAGAGAGCGCTGGAAGGGAAGG 618
Db 121875 ATGAAGAAAAAAGTCCCAAAATTTATTGAAGAGCAAGTGAAGAGAGCGCTGGAAGGGAAG 121816
Qy 619 AACAGGAGGTCCTACTTTTACGGAATTAAGCAGAGAAAAATGATGAAGAGAAAGTCAAGT 678
Db 121815 AGCAGGAGCGCCTGCTTTACAGAGTTAAAGTCGAGAAAAATGATGAAGAAAAAGTCAAGT 121756
Qy 679 TTAATTTGAGTAAGGAGCAGTAGCTCATCCGGAGCAACATCTTCCAAGTCAAGTACTC 738

```

```

Db 121755 TCAATCTGAATAAAGGAGCGGGTAGCTCAGCGGGAGCAACAGCTCCAAGTCAAGCTCGC 121696
Qy 739 TGGACCGAGTGCACCTGAAGACGATAGGAAGTTTCAAGCATCAGTGAACACGAAAAAGAACTTT 798
Db 121695 TGGACCGAGCGCACTGAAGATGCTGGGAGCACAGCATCCGTGAGGCGGAAAGAGTCTT 121636
Qy 799 CCAGAGCTCAACTCAGTCTTAAAGAAAGAAAAAGAAATCTGCACTCGATGAATCA 858
Db 121635 CACAGAGCTCAGCTCAGCCT---GCGAAGAAAGAAAGAAAGTCTGCCTCGATGAGATCA 121579
Qy 859 TGGAGATTGAAGAGGAAAAAGAAAGAACTCCCGAACAGACTACTGGCTACAGCCTCAAA 918
Db 121578 TGGAGATTGAAGAGGAAAAAGAAAGAAAGGACCGACGACAGATTTCTTGGTTACAGCTGGGA 121519
Qy 919 TTATTTGTGAAAAATTTATAACCAAGAAACTGGGAGAGAAATATCATAGAAAAAGGCTATTG 978
Db 121518 TCTTTGGGAAAAATTTATAGGAAGAAGCTTCGGGAGAAATACCAAGAAAGAGGGTCTG 121459
Qy 979 TTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTTGATCTGGAGACAAGC 1038
Db 121458 TTAAGGAAGTGAATTGACAGATACACAGCTGTGGTAAAGATGACTGACTCCGGAGACAGGC 121399
Qy 1039 TGAACCTTGACCAAGACTCTTTAGAGACAGTAAATCCAGCACCGAGGAAAAAGAAATTCAG 1098
Db 121398 TGAACCTTGGACCAAGACTCTTTAGAGACAGTCAATCCAGCCCCAGGAAAAAGAGTTCTAG 121339
Qy 1099 TTTTAAATGGAGGCTACAGAGAAATGAAGGTACCCCTAGAAATCCATCAATGAGAGAGACTT 1158
Db 121338 TTTTAAACGGAGGCTACAGAGGAAATGAAGGCAACCCCTCGAGTCCATCAATGAAAAAGACTT 121279
Qy 1159 TTTCAAGTACTGCTCAATTTGAAACTGGGCCCTTTAAAGAGCAGAGAGTTGAAGGAATTC 1218
Db 121278 TTTCAAGCAGTAGTCAATTTGAAACCGGACCTTTTGAAGGGCGCAGACTTTGAAGGTATTC 121219
Qy 1219 AATATGAAGACATTTCTTAAACTTTCCTGAGTGTGAAATTTGTTAAACAATACATTTAAAT 1278
Db 121218 AGTATGAAGACATATCTTAAACTTTCCTGAGTGTGAAATTTGTTAAACAATTTGAAT 121159
Qy 1279 CTTAAAGCATCAAAATGGTGTTCGCCAAGCATTTAGAGACTCTACTGTGTAGGGTATA 1338
Db 121158 TGTAAGCATCAAAATGTATGTCTCATGAGGTATCAAACTATGTAATTTTGTCTCTGTTA 1458
Qy 1339 TTCTTTGTATAAAAACAACAGGTTTTTGAATATTTACGTAGTTAGTTGTTACGTTAAAC 1398
Db 121098 TGTTTAT-----AAAGAAAGTCATCTATTTAAATAACCACTGGATAGCTGTTGGGTAAAT 121044
Qy 1399 TTTGAGAGAAATTTAATTTATGTCTCATGAGGTATCAAACTATGTAATTTTGTCTCTGTTA 1458
Db 121043 TTATAATAAATCTAATTTTATTTTAAAGTGCATTTTATCTTTGTTCTGTTTCTGATGG 120984
Qy 1459 TTTT 1463
Db 120983 ATTT 120979

RESULT 14
AC114615/c
LOCUS
DEFINITION Mus musculus clone RP24-82I19, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
AC114615
VERSION AC114615.5 GI:44886278
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 180882)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP24-82I19
JOURNAL Unpublished

```


REFERENCE
AUTHORS

2 (bases 1 to 180882)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bloom,I., Boguslavskiy,L.,
 Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collumore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fano,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
 Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
 MacLean,C., McDonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS

3 (bases 1 to 180882)
 Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collumore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Meldrum,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,I., Naylor,J.,
 Nguyen,C., Nicol,B., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL

Submitted (09-MAR-2004) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Mar 2, 2004 this sequence version replaced gi:24182248.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L24315

Center clone name: 82_1_19

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 191018 bases at least Q40
 Consensus quality: 191303 bases at least Q30
 Consensus quality: 191388 bases at least Q20
 Insert size: 178000; agarose-
 Insert size: 191574; sum-of-contigs
 Quality coverage: 12.2 in Q20 bases; agarose-
 Quality coverage: 11.3 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 89963: contig of 89963 bp in length
 * 89964 90063: gap of 100 bp
 * 90064 94560: contig of 4497 bp in length
 * 94561 94660: gap of 100 bp
 * 94661 169601: contig of 74941 bp in length
 * 169602 169701: gap of 100 bp
 * 169702 180882: contig of 11181 bp in length.

FEATURES

source

1..180882
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone_lib="RP24-82119"
 /clone_lib="RPC1-24 Male Mouse BAC"

misc_feature

1..89963
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 89964..90063
 /estimated_length=100
 90064..94560
 /note="assembly_fragment"
 94561..94660
 /estimated_length=100
 94661..169601
 /note="assembly_fragment"
 169602..169701
 /estimated_length=100
 169702..180882
 /note="assembly_fragment"
 clone_end:17
 vector_side:right

ORIGIN

Query Match	44.1%;	Score	673.2;	DB	14;	Length	180882;
Best Local Similarity	78.2%;	Pred. No.	1.4e-131;	Indels	6;	Gaps	1;
Matches	823;	Conservative	0;	Mismatches	223;		

QY	439	AAGGCTTGTGCAAGTGGACGACACACCAAAAGGCTGTATATTCAAGTACATAGACAGG	498
Db	46867	AAGGCTTGTGTAAGTGGATGAGACACCGAAGGCTGGTACATTCAATACATAGACAGAG	46808
QY	499	ACCCAGAAACTATCCCGCGCACTGGAACTCGAGAAAAGAAAAGCAGGACCTTGATG	558
Db	46807	ACCCAGAAACCATCCGTCGGCACTGGAATTAGAAAAAAGAAAGCAAGATCTGGACG	46748
QY	559	ATCAAGAAAAAAGTCCCAATTTATTGAGACAAAGTGAAGAGCGCTGGGAAGGAAGG	618
Db	46747	ATGAAGAAAAAAGTCCCAAGTTTCATTGAGAGAGCGTGAAGAGAGCGCTGGGAAGGAAG	46688
QY	619	AAACAGGAGGTCCTCTACTTTTACGGAAATTAAGCAGAGAAAAATGATGAAGAGAAGTCA	678
Db	46687	AGCAGGAGACCTGTGTTTTACAGAACTTTAGCCGAGAAAAATGAGGAAGAAAAAGTTAC	46628
QY	679	TTAATTTAGTAAAGAGCATGTAGCTCATCCGAGCAACATCTTCCAAGTCAAGTACTTC	738
Db	46627	TCAATCTGAATAAAGGAGCGGTGGCTCAGCGGGAGCTACAACATCCAAGTCAAGCTCTT	46568
QY	739	TGGGACCGAGTGCATCTGAAGAGCATAGGAAGTTTCAGCATCAGTGAACCGAAAAGATCTT	798
Db	46567	TGGGACCAAGTGCATCTGAAGAGCTGCTGGGAGCGGCAGCATCCCGGAAACCGAAGAGCT	46508
QY	799	CCGAGAGCTCAACTCAGTCTTAAGAAAAAGAAAAAGAAATCTGCATCTGGATGAATCA	858

Db	46507	CACAGAGCTCCGCCAGGCT-----GCCAAGAAAGAAAGTGCGCCCTGGATGAGATCA	46454
QY	859	TGGAGATTGAAGAGGAAAGAAAGAACTGCCCGAACAGACTACTCGGTACAGCCTGAAA	918
Db	46453	TGGAGCTCGAAGAGGAAAGAAAGAAAGCGCACGACGACGCGCTGGTTACAGCGGGGA	46394
QY	919	TTATTGTGAATTTATACCAAGAACTCGGGAGAGAAATATCATAAAGAAAGGCTATTG	978
Db	46393	TCGTTGTGAAATTTATACGAAGAGAGCTTGGGGAGAAATATCACAAAGAAAGGGGTCG	46334
QY	979	TTAAGGAAGTATTCACAAATATACAGCTGTGTGCAAGATGATTCATTCTGGAGCAAGC	1038
Db	46333	TTAAGGAAGTATTCACAGGTATCACAGCTGTGTGTAAGATGACTACTCTGGAGACAGC	46274
QY	1039	TGAAACTTGACAGACTCATTTAGAGACAGTAATTTCCAGCACGAGGAAAAAATTTCTAG	1098
Db	46273	TGAACCTGGACAGACTCATTTAGAGACAGTATTTCCGCCCGGGGAAAGGGTTCTAG	46214
QY	1099	TTTTAAATGGAGGCTACAGAGAAATGAAGGTACCTTGAATCCATCAATGAGAAGACTT	1158
Db	46213	TTTTAAATGGAGGCTACAGAGAAATGAAGGCACCTCTCGAATCCATCAATGAGAAGGCTT	46154
QY	1159	TTTCAGCTACTATCTCTGTAATGAACTGGCCCTTTTAAAGGAGCGAGAGTTGAAGCAATTC	1218
Db	46153	TTTCAGCCAGCATAGTCTCATTTGAACCTGGACCTTTGAAAGGAGCGAGAGTTGAAGTATTC	46094
QY	1219	AATATGAAGACATTTCTAAACTTGCTGAGTTGAAATTTGTTTAAACATACATTAAAT	1278
Db	46093	AATATGAAGACATATCTAAACTTGCTGAGTTGAAATTTGATTAACACATTTGAAAC	46034
QY	1279	CTTAAAGCATCAATTTGGTTGCGCAAGGCATTAAGACTCTACTGTTGTTAGGGTATA	1338
Db	46033	TGTGAAGCATCAATTTGGTTGAGCAAGGCACCTGTGTAACCTACTGTTGTTAGGGGATT	45974
QY	1339	TTCTTTTGTATTAACAACAACAGGTTTTTGAATAATATTACTGTATAGTTGTTTCAGCTAAC	1398
Db	45973	TGTTTTGTATTTAAAAAATAAAAAAATCATCTATTTAAATAGTAGTAAGTTGGGTAAT	45914
QY	1399	TTTGAGAAGATTTTAAATATGCTCATGAGGTATCAAACTATGTAATTTTGTCCTTGTTA	1458
Db	45913	TTATAATAATCTATGTTTTTTTAAAGTGTCATACTGTTATTTTATCTTGTTCGGTT	45854
QY	1459	TTTTTGTTCCTTTGTAATTTACTTTGATGAGT 1490	
Db	45853	TCCTGTGAGGATTTTATACCATCATTTAAAGAAAT 45822	
RESULT 15			
AC124010			
LOCUS	AC124010	182400 bp	DNA linear HTG 25-AUG-2002
DEFINITION	Mus musculus chromosome UNK clone RP23-119N4, WORKING DRAFT		
SEQUENCE	SEQUENCE, 6 unordered pieces.		
ACCESSION	AC124010		
VERSION	AC124010.2	GI:22475963	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 182400)		
AUTHORS	McPherson,J.D. and Waterston,R.H.		
TITLE	The sequence of Mus musculus clone		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 182400)		
AUTHORS	McPherson,J.D. and Waterston,R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (06-JUN-2002) Genome Sequencing Center, 4444 Forest Park		
REFERENCE	3 (bases 1 to 182400)		
AUTHORS	McPherson,J.D. and Waterston,R.H.		
TITLE	Direct Submission		

JOURNAL	Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On Aug 25, 2002 this sequence version replaced gi:21327639.
	----- Genome Center -----
	Center: Washington University Genome Sequencing Center
	Center code: WUGSC
	Web site: http://genome.wustl.edu/gsc/index.shtml
	Contact: submissions@wustl.wustl.edu
	----- Project Information -----
	Center project name: M.BA0119N04
	----- Summary Statistics -----
	Sequencing vector: M13; 0%
	Chemistry: Dye-primer ET; 0% of reads
	Chemistry: Dye-terminator Big Dye; 100% of reads
	Assembly program: Phrap; version 0.990319
	Consensus quality: 183528 bases at least Q40
	Consensus quality: 184210 bases at least Q30
	Consensus quality: 184756 bases at least Q20
	Insert size: 197000; agarose-fp
	Quality coverage: 9.53 in Q20 bases; agarose-fp
	Quality coverage: 9.61 in Q20 bases; sum-of-contigs

	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 6 contigs. The true order of the pieces
	* is not known and their order in this sequence record is
	* arbitrary. Gaps between the contigs are represented as
	* runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence
	* as soon as it is available and the accession number will
	* be preserved.

	* 1 11868: contig of 11868 bp in length
	* 11869: gap of unknown length
	* 11969: contig of 16938 bp in length
	* 28907: gap of unknown length
	* 29007: contig of 43767 bp in length
	* 72773: gap of unknown length
	* 72774: contig of 52053 bp in length
	* 124927: gap of unknown length
	* 125027: gap of unknown length
	* 125027: contig of 56876 bp in length
	* 181903: gap of unknown length
	* 182003: contig of 398 bp in length.
FEATURES	Location/Qualifiers
source	1..182400
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="UNK"
	/clones="RP23-119N4"
	1..11868
misc_feature	/note="assembly_name:Contig86"
gap	11869..11968
	/estimated_length=unknown
misc_feature	11969..28906
	/note="assembly_name:Contig87"
gap	28907..29006
	/estimated_length=unknown
misc_feature	29007..72773
	/note="assembly_name:Contig88"
gap	72774..72873
	/estimated_length=unknown
misc_feature	72874..124926
	/note="assembly_name:Contig89"
gap	124927..125026
	/estimated_length=unknown
misc_feature	125027..181902
	/note="assembly_name:Contig90"
gap	181903..182002
	/estimated_length=unknown
misc_feature	182003..182400

/note="assembly_name:Contig16"

ORIGIN

Query Match 44.1%; Score 673.2; DB 14; Length 182400;
Best Local Similarity 78.2%; Pred. No. 1.4e-131;
Matches 823; Conservative 0; Mismatches 223; Indels 6; Gaps 1;

QY 439 AAGGCTTGCAAGTGGACGAGACACCAAAAGGCTGGTATATTCAGTACATAGACAGG 498
DB 168546 AAGGCTTGTAAGTGGATGAGACACCGAAAGGCTGGTACATTCAGTACATAGACAG 168605
QY 499 ACCGAGAACTATCCGCCGCACTGGAACTGGAGAAAAGAAAAGCAGGACCTTGATG 558
DB 168606 ACCGAGAACCAATCCGTCCGCAACTGGATTAGAAAAAAGAAAGCAAGATCTGGACG 168665
QY 559 ATGAAGAAAAAATCGCAAAATTTATTGAAGAGCAAGTGAAGAGAGCCCTGGGAAGG 618
DB 168666 ATGAAGAAAAAATCGCAAGTTTCATTGAGGAGCAGGTGAGAGAGCCCTGGGAAGG 168725
QY 619 AACAGGAGTCCCTACTCTTTTACGGAATTAAGCAGAGAAATCATGAAGAGAAAGTCAGT 678
DB 168726 AGCAGGAGACACTGTTTTTACAGAACTTAGCCGAGAAATGAGGAAGAAAAGTTACGT 168785
QY 679 TTAATTTGAGTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCCAAGTCAAGTACTC 738
DB 168786 TCATCTGATTAAGGAGCGGTGGCTCAGCGGAGCTACAACATCCAAGTCAAGCTCTT 168845
QY 739 TGGGACCGAGTGCACTGAAGAGATAGGAAGTTTCAGCATCAGTGAACCGAAAAGAAATCTT 798
DB 168846 TGGGACCAAGTGCACTGAAGTGTCTGGGAGCGCAGCATCCGGGAAAACGGAAGAGTCTT 168905
QY 799 CCAGAGCTCAACTCAGTCTAAGAAAAAGAAAGAAAGAAATTCACACTGGATGAATCA 858
DB 168906 CACAGAGCTCCGCCACGCT-----GCCAAGAAAGAAAGTGGCCCTGGATGAGATCA 168959
QY 859 TGGAGATTGAAGAGGAAAAAGAAAGAACTGCCGAAACAGACTACTGGCTACAGCCCTGAAA 918
DB 168960 TGGAGCTCGAAGAGGAAAGAAAGAACCCGACGAGCAGACGCTGTTTACAGCCGGGA 169019
QY 919 TTATTGTGAAATTTATAACCAAGAAATCTGGGAGAGAAATATCATAGAAAAAGGCTATTG 978
DB 169020 TCGTTGTGAAATTTATAACCAAGAGCTTTGGGAGAAATATCACAAAGAAAGGGGTCTG 169079
QY 979 TTAAGGAAGTAAATGACAAATATACAGCTGTTGTGAAAGATGATTGATCTGGAGACAAGC 1038
DB 169080 TTAAGGAAGTAAATGACAGTACACAGCTGTGGTAAAGATGACTGACTCTGGAGACAGC 169139
QY 1039 TGAACCTTGACAGACTCATTTTAGACACAGTAAATCCAGCACCAGGAAAAAGAAATCTTAG 1098
DB 169140 TGAACCTTGGACAGACTCATTTTAGACACAGTCAATCCGGCCCGGGGAAAAGGGTTCTAG 169199
QY 1099 TTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCTTAGAATCCATCAATGAGAAGACTT 1158
DB 169200 TTTTAAATGGAGGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCAATGAGAAGCTT 169259
QY 1159 TTTTCAGCTACTATCGTCAATTGAAACTTGGCCCTTTTAAAGGACCGCAGATTGAAGGAATTC 1218
DB 169260 TTTTCAGCCACGATAGTCAATTGAAACTTGGACCTTTGAAAGGACGACAGTTGAAGGTATTTC 169319
QY 1219 AATATGAAGACATTTCTAAACTTGCCTGAGTTTGAATAATTTGTTAAACAATACATAAAT 1278
DB 169320 AATATGAAGACATATCTAAACTTGCCTGAGTTTGAATAATTTGTTAAACAATACATAA 169379
QY 1279 CTTTAAAGCATCAATTTGGTGTCCCAAGGCATTTATGAGACTCTACTGTGTAGGGTATA 1338
DB 169380 TGTGAAGCATCAATTTGGTGTAGCAAGGCACTGTGTAACTCTACTGTGTAGGGGATT 169439
QY 1339 TCTTTTGTATAAAACAAACAGGTTTTTGAATAATTTACTGTATAGTTGTTTCAAGCTAAAC 1398
DB 169440 TGTTTTGTATAAAAAAATAAATCATCTATTATAATACTAGTGAATAGTTGGGTAAAT 169499
QY 1399 TTTGAGAGAAATTTAATATGCTCATGAGGATCAAACTATGTAATTTTGTCTCTTGTTA 1458

Db 169500 TTATAATAAAATCTATGTTTTTTTTTAAGTGTCAATACTGTTATTTATCTTTGTTCCGTT 169559
QY 1459 TTTTGTGTTCCCTTTGTAAATTTTACTTTGATGACT 1490
Db 169560 TCCCTGATGGATTTTACCATCATTTAAGAAT 169591

Search completed: November 26, 2005, 19:10:38
Job time : 8131.42 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 980.087 Seconds
(without alignment)
10390.557 Million cell updates/sec

Title: US-09-555-529-1
Perfect score: 1528
Sequence: 1 ctgaatcagcgccgctg.....atgtattataaaaaaaaaa 1528

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1249.8	81.8	1296	2	AAX85549
2	1002	65.6	1002	2	AAX85552
3	990.6	64.8	1390	2	AAX85570
4	986.4	64.6	1458	2	AAQ79936
5	653.4	42.8	1002	2	AAX85551
6	531.6	34.8	679	13	ADQ56943
7	490.8	32.1	591	4	AAX85550
8	475	31.1	1102	2	AAX85550
9	380.6	24.9	1372	12	ADP22451
10	303.8	19.9	398	13	ACF87901
11	298.6	19.5	1241	4	ABL29199
12	284	18.6	300	3	AAA01559
13	247.2	16.2	3313	4	ABL29198
14	224.4	14.7	1459	3	AAC47248
15	209.4	13.7	4394	4	ABL29184
16	209.4	13.7	4582	4	ABL18194
17	204.8	13.4	1713	13	ADT17629
18	204.8	13.4	1780	3	AAA52591
19	204.8	13.4	1780	3	AAA52590

20	201.6	13.2	1614	13	AD084255
21	201.6	13.2	1713	13	ADx47874
22	201.6	13.2	1764	3	AAx52589
23	136.4	8.9	1032	3	AAc48794
24	126.4	8.3	549	3	AAc94929
25	100.2	6.6	270	6	ABL73715
26	94.2	6.2	696	3	AAc38477
27	77.4	5.1	431	3	AAc48801
28	60	3.9	60	6	ABn33592
29	59.4	3.9	12578	4	AAa46660
30	58	3.8	5935	4	AAa45427
31	56.8	3.7	2985	4	ABL29210
32	55.2	3.6	11422	6	ABK39937
33	55.2	3.6	11422	6	ABL32219
34	55	3.6	2000	8	ADA71938
35	54.4	3.6	65	6	ABn51894
36	54	3.5	15951	6	ABn33680
37	54	3.5	15951	6	ABL34580
38	54	3.5	15951	6	ABL70373
39	54	3.5	15951	7	ADs99841
40	53.8	3.5	6831	6	ABL33486
41	53.4	3.5	6013	6	ABK31360
42	53.4	3.5	6013	6	AAa61264
43	53	3.5	8056	8	ABZ10246
44	52.6	3.4	2000	11	ACL37108
45	52.6	3.4	113515	6	ABL34175

ALIGNMENTS

RESULT 1
AAX85549
ID AAX85549 standard; cDNA; 1296 BP.
XX
AC AAX85549;
XX
DT 07-SEP-1999 (first entry)
XX
DE cDNA of a gene coding for the human kin17 protein.
XX
KW Human; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme; ss.
XX
OS Homo sapiens.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PA Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX
PI WPI; 1999-359999/31.
XX
DR New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
PT Claim 1; Page 30-31; 69pp; French.
XX
PS The present sequence encodes a human kin17 protein. The mammalian kin17
XX protein is useful for preparing a medicament for controlling cell
XX proliferation or for controlling fertility. The medicaments can also be
XX used to treat hyperproliferative diseases. Fragments between amino acids
XX 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
XX kin17 protein are useful for regulating the interaction between proteins
XX and curved DNA. The fragment can be used to block replication of HIV or
XX and curved DNA.

CC its integration into the human genome or to target repair enzymes to
CC curved DNA sites. Expression vectors for kin17 can be used for
CC controlling cell proliferation

SQ Sequence 1296 BP; 453 A; 240 C; 310 G; 293 T; 0 U; 0 Other;

Query Match 81.8%; Score 1249.8; DB 2; Length 1296;
Best Local Similarity 99.8%; Pred. No. 7.1e-292;
Matches 1262; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY	40	AGAAAGTATCGCTCGCGTGGTGGCATGGGAAGTCGGATTTCTTCTCCCAAGGCTA	99
DB	32	AGAAAGTATCGCTCGCGTGGTGGCATGGGAAGTCGGATTTCTTCTCCCAAGGCTA	91
QY	100	TCGCCAACAGGATCAAGTCCAAAGGGGCTGCAGAAGCTACGCTGGTATTGCCAGATGTGCC	159
DB	92	TCGCCAACAGGATCAAGTCCAAAGGGGCTGCAGAAGCTACGCTGGTATTGCCAGATGTGCC	151
QY	160	AGAAGCAGTCCGGGACGAGAAATGGCTTTAAGTGCATTTATGTCCGAATCTCATCAGA	219
DB	152	AGAAGCAGTCCGGGACGAGAAATGGCTTTAAGTGCATTTATGTCCGAATCTCATCAGA	211
QY	220	GACAACTATTGCTGCTTCAGAAAATCCTCAGCAGTTTATGGATTATTTTTCAGAGGAAT	279
DB	212	GACAACTATTGCTGCTTCAGAAAATCCTCAGCAGTTTATGGATTATTTTTCAGAGGAAT	271
QY	280	TCCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAACA	339
DB	272	TCCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAACA	331
QY	340	ACATTGTCTAACAGAAATACATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGT	399
DB	332	ACATTGTCTAACAGAAATACATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGT	391
QY	400	GGGAACTCTGACTGATTTTACTAGTGGCTGGGACAGAGAGCTTGTGCAAGTGGACG	459
DB	392	GGGAACTCTGACTGATTTTACTAGTGGCTGGGACAGAGAGCTTGTGCAAGTGGACG	451
QY	460	AGACACCAAAGGCTGGTATTTTACGTACATAGACAGGAGCCAGAAAATATCCGCCCGC	519
DB	452	AGACACCAAAGGCTGGTATTTTACGTACATAGACAGGAGCCAGAAAATATCCGCCCGC	511
QY	520	AACTGGAATCGAGAAAAAGAAAAGCAGGACCTTGATGATGAAGAAAATCTGCCAAAT	579
DB	512	AACTGGAATCGAGAAAAAGAAAAGCAGGACCTTGATGATGAAGAAAATCTGCCAAAT	571
QY	580	TTATTGAAGAGCAAGTGAAGAGAGCGCTTGGAGGGAAGCAAGCAGGAGTCCCTACTTTTA	639
DB	572	TTATTGAAGAGCAAGTGAAGAGAGCGCTTGGAGGGAAGCAAGCAGGAGTCCCTACTTTTA	631
QY	640	CGGAATTAAGCAGAGAAAAATGATGAAGAGAAAGTCAACGTTTAAATTTGATGAAGAGCAT	699
DB	632	CGGAATTAAGCAGAGAAAAATGATGAAGAGAAAGTCAACGTTTAAATTTGATGAAGAGCAT	691
QY	700	GTAGTCTATCCGGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGACATGAAGA	759
DB	692	GTAGTCTATCCGGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGACATGAAGA	751
QY	760	CGATAGGAAGTTTCAGCATCAGTGAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTA	819
DB	752	CGATAGGAAGTTTCAGCATCAGTGAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTA	811
QY	820	AAGAAAAGAGAAAAAAGAAATCTGCATCGATGAATCATGGAGATTTGAAGAGGAAAAAGA	879
DB	812	AAGAAAAGAGAAAAAAGAAATCTGCATCGATGAATCATGGAGATTTGAAGAGGAAAAAGA	871
QY	880	AAAGAACTGCCGACAGACTACTGGCTACAGCTGAAATTTATGTGAAAATTAACCA	939
DB	872	AAAGAACTGCCGACAGACTACTGGCTACAGCTGAAATTTATGTGAAAATTAACCA	931
QY	940	AGAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAAGTAAATTGACAAT	999
DB	932	AGAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAAGTAAATTGACAAT	991

QY	1000	ATACAGCTGTGTGAAGATGATTGATTTCTGGAGACAAAGCTGAAACTTGGACAGACTCATT	1059
DB	992	ATACAGCTGTGTGAAGATGATTGATTTCTGGAGACAAAGCTGAAACTTGGACAGACTCATT	1051
QY	1060	TAGAGACAGTAATTTCCAGACCCAGGAAAAGAAATCTAGTTTAAATGGAGGCTACAGAG	1119
DB	1052	TAGAGACAGTAATTTCCAGACCCAGGAAAAGAAATCTAGTTTAAATGGAGGCTACAGAG	1111
QY	1120	GAATGAAGGTACCCCTAGATCCATCAATGAGAAGACTTTTTCAGCTACTATCGTCATTG	1179
DB	1112	GAATGAAGGTACCCCTAGATCCATCAATGAGAAGACTTTTTCAGCTACTATCGTCATTG	1171
QY	1180	AAACTGGCCCTTTAAAAGGACGACAGATTGAAGGAATTCATATGAAGACATTTCTAAAC	1239
DB	1172	AAACTGGCCCTTTAAAAGGACGACAGATTGAAGGAATTCATATGAAGACATTTCTAAAC	1231
QY	1240	TTGCTGAGTTTGAATAATTTGTTTAAACAATA-CATTAAATCTTAAAGCATCAAAATTTGGTG	1298
DB	1232	TTGCTGAGTTTGAATAATTTGTTTAAACAATACCTTTAAATCTTAAAGCATCAAAATTTGGTG	1291
QY	1299	TTGCG 1303	
DB	1292	TTGCG 1296	

RESULT 2

AA85552
ID AA85552 standard; cDNA; 1002 BP.

AC AA85552;

DT 07-SEP-1999 (first entry)

DE Probe derived from cDNA of a gene coding for the human kin17 protein.

DE Human; kin17 protein; cell proliferation; fertility;

KW hyperproliferative disease; protein interaction; curved DNA;

KW HIV replication; HIV integration; repair enzyme; probe; ss.

OS Homo sapiens.

PN FR2772046-A1.

PD 11-JUN-1999.

PF 09-DEC-1997; 97FR-00015536.

PR 09-DEC-1997; 97FR-00015536.

PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

DR WPI; 1999-359999/31.

PT New DNA coding for human kin17 protein - useful for controlling cell

PS proliferation or fertility.

CC Claim 7; Page 32-33; 69pp; French.

CC The present sequence is derived from nucleotides 207-1208 of AA85549,
CC and is a probe used to isolate human kin17 cDNA. The mammalian kin17
CC protein is useful for preparing a medicament for controlling cell
CC proliferation or for controlling fertility. The medicaments can also be
CC used to treat hyperproliferative diseases. Fragments between amino acids
CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
CC kin17 protein are useful for regulating the interaction between proteins
CC and curved DNA. The fragment can be used to block replication of HIV or
CC its integration into the human genome or to target repair enzymes to
CC curved DNA sites. Expression vectors for kin17 can be used for
CC controlling cell proliferation

SQ	Sequence 1002 BP; 374 A; 179 C; 235 G; 214 T; 0 U; 0 Other;	
Query Match	65.6%; Score 1002; DB 2; Length 1002;	
Best Local Similarity	100.0%; Pred. No. 5.6e-232;	
Matches 1002; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	215 TCAGAGACAACATATTGCTGGCTTCAGAAAAATCCTCAGCAGCTTTATGGATTATTTTTCAGA 274	
DB	1 TCAGAGACAACATATTGCTGGCTTCAGAAAAATCCTCAGCAGCTTTATGGATTATTTTTCAGA 60	
QY	275 GGAATTCGGAATGACTTTCTAGAACTTTCTCAGGAGACGCTTTGGCACTAAAGGGTCCA 334	
DB	61 GGAATTCGGAATGACTTTCTAGAACTTTCTCAGGAGACGCTTTGGCACTAAAGGGTCCA 120	
QY	335 CAACAACATTGCTTACAAGCAATATCATCAGCCACCGAGAGCACTCCACATGAATGCCAC 394	
DB	121 CAACAACATTGCTTACAAGCAATATCATCAGCCACCGAGAGCACTCCACATGAATGCCAC 180	
QY	395 TCAGTGGGAACTCTGACTGATTTTACTAAGTGGCTGGCGAGAGAGGCTTGTGCAAGT 454	
DB	181 TCAGTGGGAACTCTGACTGATTTTACTAAGTGGCTGGCGAGAGAGGCTTGTGCAAGT 240	
QY	455 GGACGAGACACCAAAAGGCTGTATATTACGTACATACAGAGGACCCAGAACTATCCG 514	
DB	241 GGACGAGACACCAAAAGGCTGTATATTACGTACATACAGAGGACCCAGAACTATCCG 300	
QY	515 CCGCAACTGGAACTGGAGAAAAAGAAAGCAGACCTTGATGATGAAGAAAAAATCTGC 574	
DB	301 CCGCAACTGGAACTGGAGAAAAAGAAAGCAGACCTTGATGATGAAGAAAAAATCTGC 360	
QY	575 CAATTTTATGAAGACCAAGTCAGAGAGGCTTGAAGGGAAGGAGGAGGCTCCCTAC 634	
DB	361 CAATTTTATGAAGACCAAGTCAGAGAGGCTTGAAGGGAAGGAGGAGGCTCCCTAC 420	
QY	635 TTTTACGGAATTAAGCAGAGAAAAATGATGAAGAGAAAGTCAGCTTTAATTTGAGTAAAGG 694	
DB	421 TTTTACGGAATTAAGCAGAGAAAAATGATGAAGAGAAAGTCAGCTTTAATTTGAGTAAAGG 480	
QY	695 AGCATGTAGTCACTCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGAGTGCAT 754	
DB	481 AGCATGTAGTCACTCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGAGTGCAT 540	
QY	755 GAAGACGATAGCACTGAGCACTAGTGAAGAAAGAAAGAAATCTCCAGAGCTCAACTCA 814	
DB	541 GAAGACGATAGCACTGAGCACTAGTGAAGAAAGAAAGAAATCTCCAGAGCTCAACTCA 600	
QY	815 GTCTAAAGAAAAAGAAAAAGAAATCTGCATCTGGATGAATCATGGAGATTGAAGAGGA 874	
DB	601 GTCTAAAGAAAAAGAAAAAGAAATCTGCATCTGGATGAATCATGGAGATTGAAGAGGA 660	
QY	875 AAAGAAAAAGAAATGCGCGAAGCACTACTGGCTACAGCTCGAAATTAATTTGTAATAATTA 934	
DB	661 AAAGAAAAAGAAATGCGCGAAGCACTACTGGCTACAGCTCGAAATTAATTTGTAATAATTA 720	
QY	935 AACCAAGAACTGGGAGAGAAATATCATGAAGAAAGGCTATTGTTAAGAAAGTAAATGA 994	
DB	721 AACCAAGAACTGGGAGAGAAATATCATGAAGAAAGGCTATTGTTAAGAAAGTAAATGA 780	
QY	995 CAAATATACAGCTGTTGTGAAGATGATGATTTCTGGAGACAAGCTGAAACTTGACCAGAC 1054	
DB	781 CAAATATACAGCTGTTGTGAAGATGATGATTTCTGGAGACAAGCTGAAACTTGACCAGAC 840	
QY	1055 TCATTTAGAGACAGTAATTTCCAGACACAGAAAAAGAAATTTAGTTTTTAAATGGAGGCTA 1114	
DB	841 TCATTTAGAGACAGTAATTTCCAGACACAGAAAAAGAAATTTAGTTTTTAAATGGAGGCTA 900	
QY	1115 CAGAGAAATGAGGTACCTTAGAATCCATCATGAGAGACCTTTTTCAGCTACTATCGT 1174	
DB	901 CAGAGAAATGAGGTACCTTAGAATCCATCATGAGAGACCTTTTTCAGCTACTATCGT 960	
QY	1175 CATTGAAACTGGCCCTTTAAAGGACCGCAGAGTTGAAGGAAT 1216	
DB	961 CATTGAAACTGGCCCTTTAAAGGACCGCAGAGTTGAAGGAAT 1002	

RESULT 3

AAx85570
ID AAx85570 standard; cDNA; 1390 BP.

XX
AC AAx85570;

XX
DT 07-SEP-1999 (first entry)

XX
DE cDNA of a gene coding for the murine kin17 protein.

XX
KW Mouse; kin17 protein; cell proliferation; fertility;

XX
KW hyperproliferative disease; protein interaction; curved DNA;

XX
KW HIV replication; HIV integration; repair enzyme; ss.

XX
OS Mus sp.

XX
PN FR2772046-A1.

XX
PD 11-JUN-1999.

XX
PF 09-DEC-1997; 97FR-00015536.

XX
PR 09-DEC-1997; 97FR-00015536.

XX
PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX
PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJP;

XX
PS WPI; 1999-359999/31.

XX
PT New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.

XX
PS Claim 21; Page 35-36; 69pp; French.

XX
CC The present sequence encodes a murine kin17 protein. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation

XX
SQ Sequence 1390 BP; 482 A; 256 C; 353 G; 299 T; 0 U; 0 Other;

Query Match 64.8%; Score 990.6; DB 2; Length 1390;

Best Local Similarity 83.0%; Pred. No. 3.7e-229; Indels 6; Gaps 1;

Matches 1144; Conservative 0; Mismatches 229;

QY 66 ATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAAAGGG 125

DB 1 ATGGGCAAGTCGGATTTTCTGAGCCCAAGGCCATAGAAATTAAGTCCAAAGGG 60

QY 126 CTGCAGAAAGCTACGCTGGTATTTGCCAGATGTGCCAGAGCAGTGCCTGGGACGAGATGGC 185

DB 61 CTCCAGAAAGCTTCGCTGGTACTGTCCAGATGTGCCAAAAGCAATGCCGACGAGATGGC 120

QY 186 TTTAAGTGTCTATGTATGTCGCAATCTCATCAGAGCAACTATTTGCTGCTTCAGAAAAAT 245

DB 121 TTTAAGTGTCTATGTATGTCGCAATCTCATCAGAGCAACTATTTGCTGCTTCAGAAAAAT 180

QY 246 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTTCCGAAATGACTTTTCTAGAACTTCTC 305

DB 181 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTTCCGAAATGACTTTTCTGGAACCTTG 240

QY 306 AGGAGACGCTTTGGCACTAAAGGGTCCACAAACATTTGTCTACAAACATATCATCAGC 365

DB 241 AGGCGACGCTTTGGCACTAAAGGGTCCACAAACATTTGTCTACAAATGAATATCATCAGC 300

Qy 366 CACGAGAGCATCCACATGATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAG 425
Db 301 CACGAGAGCATCCACATGATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAG 360
Qy 426 TGGCTGGGACAGAGGCTTGTGCAAGTGCAGACACACCAAAAGGCTGTATATTACG 485
Db 361 TGGCTGGGACAGAGGCTTGTGCAAGTGCAGACACACCAAAAGGCTGTATATTACG 420
Qy 486 TACATAGACAGGACCCAGAAATATCCCGCGCAACTTGGAACTGGAGAAAAAGAAAAAG 545
Db 421 TACATAGACAGACCCAGAAATATCCCGCGCAACTTGGAACTGGAGAAAAAGAAAAAG 480
Qy 546 CAGACCTTGATGATGAAGAAAAAATCTCCAAAATTTATGAAGAGCAAGTGAAGAGGC 605
Db 481 CAAGATCTGGACGATGAAGAAAAAATCTCCAAAATTTATGAAGAGCAAGTGAAGAGGC 540
Qy 606 CTGGAAGGGAAGGACAGGAGTCCCTACTTTTACGGAATTAAGCAGAGAAATGATGA 665
Db 541 CTGGAAGGGAAGGACAGGAGTCCCTACTTTTACGGAATTAAGCAGAGAAATGATGA 600
Qy 666 GAGAAAGTCAGCTTTAAATTTGAGTAAAGAGCATGTAGCTCATCCGAGCAACATCTTCC 725
Db 601 GAAAAAGTTACGTTCAATCTGAATAAAGAGCGGTGCTCAGCGGAGCTTACAACATCC 660
Qy 726 AAGTCAAGTACTCTGGGACCGAGTGCATGAAGAGCATAGGAAGTTTCAAGATCAGTGAA 785
Db 661 AAGTCAAGTCTTTGGGACCAAGTGCATGAAGTGTCTGGGAGCGCAGCATCCGGGAAA 720
Qy 786 CGAAAAAGATCTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAGAAAAAGAAATCTGCA 845
Db 721 CGAAAAAGATCTCCAGAGCTCCCGCCAGCTGC-----GAAGAAGAAAGTCCGCC 774
Qy 846 CTGGATGAATCATGGAGATTGAAGAGGAAAAAGAACTGCCCGGAGCACTACTCGG 905
Db 775 CTGGATGAGATCATGGAGCTCGAAGAGGAAAGAAAGGACCGCACGAGACGCTCGG 834
Qy 906 CTACAGCTCGAAATTTATGTGAAATTTAATCAAGAAATCTGGAGAGAGAAATATCATAG 965
Db 835 TTACAGCCGGGTCGTTGTGAAATTTAATCAAGAAAGTGGGAGAGAAATATCACAAG 894
Qy 966 AAAAAGGCTATTGTGTAAGAGTAAATTCACAAATATACAGCTGTTGTAAGATGATTGAT 1025
Db 895 AAAAAAGGCTCGTTTAAGAAAGTGAATGACAGGTACACAGCTGTGTTAAAGATGACTAC 954
Qy 1026 TCTGGAGACAAGCTGAACTTTGACCAGACTCATTTAGAGACAGTAATTTCCAGCACGGA 1085
Db 955 TCTGGAGACAGCTGAACTGGACCAGACTCATTTAGAGACAGTCAATTCGCGCCCGGG 1014
Qy 1086 AAAAGAATCTAGTTTTAAATGGAGGCTACAGAGAAATGAAGTACCCTAGAAATCCATC 1145
Db 1015 AAAAGGGTCTAGTTTTAAATGGAGGCTACAGAGAAATGAAGGACCTCTCGAATCCATC 1074
Qy 1146 AATCAGAGAGCTTTTTCAGCTACTATCGTCAATGAACTGGCCCTTTTAAAGAGGACGAGA 1205
Db 1075 AATCAGAGAGGCTTTTTCAGCCACAGATGATCAATGAACTGGACCTTTGAAAGGAGCGAGA 1134
Qy 1206 GTTGAAGGAATTCATATGAAGACATTTTCTAACTTGCTGAGTTTGAAAAATTTGTTAAC 1265
Db 1135 GTTGAAGGATTCATATGAAGACATATCTAACTTGCTGAGTTTGAAAAATTTGATAAC 1194
Qy 1266 AATACATTAATCTTAAGGATCAAAATGGTGTTCGCCAGGCGATATGAGACTCTACT 1325
Db 1195 AACACATTAAGGATCAAAATGGTGTTCGCCAGGCGATATGAGACTCTACT 1254
Qy 1326 GTGTTAGGGTATTTCTTTTGTATAAAACAAACAGGTTTTTGAATAATTTACTGTATAGT 1385
Db 1255 GTGTTAGGGATTTGTTTGTATTAAAAAAAATAATCATCTATTTAAATACTAGTGA 1314
Qy 1386 TGTTCAAGTAACTTTGAGAGAAATTTTAAATATGTTCTCATGAGGTATCAAACTATGTA 1444
Db 1315 TAGTTGGGTAAATTTATATAAATCTATGTTTTTTTAAAGTGTATAAAAAAATAAAAAA 1373

RESULT 4
AAQ79936
ID AAQ79936 standard; cDNA; 1458 BP.
XX AC AAQ79936;
XX AC AAQ79936;
DT 25-MAR-2003 (revised)
DT 06-SEP-1995 (first entry)
XX Murine Kin17 cDNA.
XX chromosomal rearrangement; kin17 protein; SOS DNA repair system; RecA;
KW genotoxic agent; zinc finger; DNA binding protein; ss.
XX Mus musculus.
XX Key Location/Qualifiers
FT primer_bind complement(1..21)
FT /tag= b
FT /note= "Oligo L (AAQ79946) binds to complementary strand"
FT misc_feature 22..1434
FT /tag= 1
FT /label= kin17 cDNA
FT /note= "nucleotides 1-1414; the genomic DNA contains at
FT least two introns within this sequence, see Comments"
FT primer_bind 32..49
FT /tag= c
FT /note= "Oligo C (AAQ79938) binding site"
FT CDS 46..1221
FT /tag= a
FT /product= "Kin17"
FT /note= "N's in the sequence denote illegible residues"
FT primer_bind complement(67..86)
FT /tag= d
FT /note= "Oligo S (AAQ79947) binds to complementary strand"
FT primer_bind 274..297
FT /tag= e
FT /note= "Oligo D (AAQ79939) binding site"
FT primer_bind complement(339..360)
FT /tag= f
FT /note= "Oligo K (AAQ79945) binds to complementary strand"
FT primer_bind 451..474
FT /tag= g
FT /note= "Oligo J (AAQ79944) binding site"
FT primer_bind complement(550..567)
FT /tag= h
FT /note= "Oligo E (AAQ79940) binds to complementary strand"
FT primer_bind 802..825
FT /tag= i
FT /note= "Oligo F (AAQ79941) binding site"
FT primer_bind complement(839..862)
FT /tag= j
FT /note= "Oligo G (AAQ79942) binds to complementary strand"
FT primer_bind complement(1435..1458)
FT /tag= k
FT /note= "Oligo B (AAQ79937) binds to complementary strand"
XX FR2706487-A1.
PN 23-DEC-1994.
XX 23-DEC-1994.
XX 15-JUN-1993; 93FR-00007171.
XX 15-JUN-1993; 93FR-00007171.
XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX Angulo-Mora JF, Tissier A, Frelat G, Mauffrey P, Guilly M;
XX WPI; 1995-039031/06.
XX Purified murine kin17 protein prepn. for detecting chromosomal

PT rearrangements - also related antibodies, human and murine DNA, primers,
 XX probes and vectors, used to assess damage caused by genotoxic agents.
 PS Claim 9; Page 33; 54pp; French.
 XX The murine Kin17 protein includes a zinc finger domain (see AAR66766),
 CC recognises single- and double-stranded DNA (partic. regions of secondary
 CC structure), has apparent mol. wt. 43 kD and is recognised by both anti-
 CC kin17 antibodies and antibodies against the RecA protein of E.coli. The
 CC kin17 protein is involved in DNA repair; it can be used to monitor
 CC chromosomal rearrangements following exposure to genotoxic agents. The
 CC kin17 cDNA sequence AA079936 consists of a 1414 nucleotide sequence,
 CC flanked by primer binding sites; the genomic kin17 DNA is claimed in
 CC which an intron is inserted at position 137 of the 1414 nucleotide cDNA
 CC and a second intron is located between nucleotides 339-429 of the cDNA.
 XX (Updated on 25-MAR-2003 to correct PN field.)
 SQ Sequence 1458 BP; 496 A; 272 C; 372 G; 311 T; 0 U; 7 Other;

Query Match 64.6%; Score 986.4; DB 2; Length 1458;
 Best Local Similarity 81.9%; Pred. No. 3.9e-228;
 Matches 1147; Conservative 0; Mismatches 248; Indels 6; Gaps 1;

QY 44 AGTGATCGCTGCGTGGTCCCATGGGAGTGGGATTTCTTACTCCCAAGGCTATCGC 103
 DB 24 AGTTGTAGCGCTCAGGCTCGCATGGGCAANTCGGATTTCTGAGCCCNAGGCCATCGC 83
 QY 104 CAACAGGATCAAGTCCCAAGGGGCTCGAGAACTACGCTGGTATGCCAGATGTGCCAGAA 163
 DB 84 CAATAGAAATTAAGTCCAAAGGGCTCCAGAAACGTTNNNTNNTACTGCCAGATGTGCCAAA 143
 QY 164 GAAGTCCGGGACGAGAAATGGCTTTTAAGTGTCAATGTATGTCCGATCTCATCAGAGACA 223
 DB 144 GCAATGCCCGACGAGAAATGGCTTTTAAGTGTCACTGTATGTCTGAATCTCATCAAGACA 203
 QY 224 ACTATTGCTGGCTTCAGAAATCTCCAGAGTCTTATGGATTTATTTTCAGAGAAATTCGG 283
 DB 204 ACTGTTGCTGGCTTCAGAAATCTCCAGAGTCTTATGGATTTATTTTCAGAGAAATTCGG 263
 QY 284 AAATGACTTTCTAGAACTTTCTCAGGAGACGCTTTGGCACTTAAAGGGTCCCAACAACAT 343
 DB 264 AAATGACTTTCTGGAATCTTCAGGGCGCGCTTTGGCACTTAAAGGGTCCCAACAACAT 323
 QY 344 TGTCTCAACAGATATACATCAGCCACCGAGAGCNCATCCACATGAATGCCACTCAGTGGGA 403
 DB 324 TGTCTCAACATGAATACATCAGCCACCGAGAGCNCATCCACATGAATGCCACTCAGTGGGA 383
 QY 404 AACTCTGACTGATTTTACTTAAGTGGCTGGCGAGAGAGGCTTGTGCAAAAGTGGACGAGAC 463
 DB 384 GACACTGACCGACTTTTACCAGTGGCTGGCGAGAGAGGCTTGTGTAAGTGGATGAGAC 443
 QY 464 ACCAAAGGCTGCTATATTTCAGTACATAGACAGGACCCAGAACTATCCGCGGCAACT 523
 DB 444 ACCGAAGGCTGCTATATTTCAGTACATAGACAGGACCCAGAACTATCCGCTGGCAACT 503
 QY 524 GGAATCTGGAGAAAGAAAGCAGACCTTGATGATGAAGAAAGAACTGCCAAATTTAT 583
 DB 504 GGAATTTAGAAAGAAAGAAAGCAGACCTTGATGATGAAGAAAGAACTGCCAAATTTAT 563
 QY 584 TGAAGAGCAGTTCAGAGAGGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 643
 DB 564 TGAAGAGCAGTTCAGAGAGGCTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 623
 QY 644 ATTAAGCAGAGAAATATGATGAAGAGAAAGTCAAGTTTAAATTTTGAAGTAAAGGAGCATGTAG 703
 DB 624 ACTTAGCCGAGAAATATGAGGAGAAAGAAAGTTTACGTTCAATCTGAATTAAGGAGCGGTGG 683
 QY 704 CTCAATCCGGAGCAACATCTTTCCAAAGTCAAGTACTCTGGGACCCAGTGCATGAAGAGCAT 763
 DB 684 CTCAAGGGGAGCTACAACTCCAAAGTCAAGTCTTTGGGACCAAGTGCATGAAGTGTCT 743
 QY 764 AGGAAGTTTCAGCATCAGTGAAGCAAGAAAGATCTTCCAGAGCTCAACTCAGTCTAAAGA 823

744 GGGGAGCGCAGCATCCGGGAAACGGAAGAGTCTTTACAGAGCTCCGCCAGCCT----- 798
 QY 824 AAAGAGAGAAAGAAATCTGCACCTGGATGAATCATGAGAGATTTGAAGAGGAAAGAAAG 883
 DB 799 -GCGAAGAGAAAGAGTCCGCCCTGGATGAGATCATGAGCTCGAAGAGGAAAGAAAG 857
 QY 884 AACTGCCCGAAGACAGACTACTGGCTACAGCCTGAAATTTATTTGAAAAATTTATAACCAAGAA 943
 DB 858 GACCGACGGACAGACGGCTGTTTACAGCCGGGATCGTTGTGAAAAATTTATAACCAAGAA 917
 QY 944 ACTGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGAGTAAATTTGCAAAATATATAC 1003
 DB 918 GCTTGGGAGAAATATCATAGAAAAAGGCTCGTTAAGGAAGTGAATGACAGGTACAC 977
 QY 1004 AGCTGTTCTGAGAGTGAATTTGATTTCTGGAGACAAAGCTGAAACTTTGACACAGACTCATTTAGA 1063
 DB 978 AGCTGTGTAAGATGACTGCTGGAGACAGGCTGAAACTGGACACAGACTCATTTAGA 1037
 QY 1064 GACAGTAATTTCCAGCACCGAGAAAGAAATTTCTAGTTTAAATGGAGGCTACAGAGGAAA 1123
 DB 1038 GACAGTCAATTCGGGCCCGGGGAAAGGGTCTTAGTTTAAATGGAGGCTACAGAGGAAA 1097
 QY 1124 TGAAGGTACCTTAGAATCCATCAATGAGAAGACTTTTTTCAGCTACTATCTCGTCAATTGAAAC 1183
 DB 1098 TGAAGGCACCTCTCGAATCCATCAATGAGAAGCTTTTTTCAGCCACGATAGTCAATTGAAAC 1157
 QY 1184 TGGCCCTTTAAAGGACGACAGAGTTGAAAGAAATTCATATATGAAGACATTTCTAAACTTGC 1243
 DB 1158 TGGACCTTTGAAAGGACGACAGTTGAAAGTATTCATATATGAAGACATATCTTAACTTGC 1217
 QY 1244 CTGAGTTTGAATAATTTGTTTAAACAATACATTTAAATCTTTAAAGCATCAAATTCGTGTTCG 1303
 DB 1218 TTGAGTTTGAATAATTTGATTAACAACATTTGAAACTGTGAAGCATCAAATTCGTGTTCG 1277
 QY 1304 CAAGCATATGAGACTCTACTGTAGGGTATATCTTTTGTATATAAACAACACAGGTT 1363
 DB 1278 CAAGGCACCTGTGAATCTCTACTGTAGGGGATTTGTTTGTATTTAAAAAAA 1337
 QY 1364 TTTGAAATATTTACTGTATAGTTGTTTCACTAAACTTTTGAAGAAATTTTAAATTTATGTCCTC 1423
 DB 1338 CATCTATTAAATATCTAGTGAATAGTTGGTAAATTTTAAATTTATATATCTATGTTTTTTT 1397
 QY 1424 ATGAGGTATCAAACTATGTAA 1444
 DB 1398 AAGTGTAAAAA 1418

RESULT 5
 AAX85551
 ID AAX85551 standard; cDNA; 1002 BP.
 XX AAX85551;
 AC AAX85551;
 XX 07-SEP-1999 (first entry)
 DT 07-SEP-1999
 XX cDNA of a gene coding for a truncated human kin17 protein.
 DE kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 XX protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme; ss.
 XX Homo sapiens.
 OS PR2772046-A1.
 XX 11-JUN-1999.
 PD 09-DEC-1997; 97FR-00015536.
 PF 09-DEC-1997; 97FR-00015536.
 XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 PA

PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
DR
XX New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.
XX
XX Claim 5; Page 32; 69pp; French.
XX
XX The present sequence encodes a truncated human kin17 protein with amino
CC acids 129-228 deleted. The mammalian kin17 protein is useful for
CC preparing a medicament for controlling cell proliferation or for
CC controlling fertility. The medicaments can also be used to treat
CC hyperproliferative diseases. Fragments between amino acids 55 and 235
CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein
CC are useful for regulating the interaction between proteins and curved
CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
XX proliferation
SQ Sequence 1002 BP; 343 A; 189 C; 230 G; 240 T; 0 U; 0 Other;
Query Match 42.8%; Score 653.4; DB 2; Length 1002;
Best Local Similarity 76.6%; Pred. No. 1.2e-147;
Matches 969; Conservative 0; Mismatches 1; Indels 295; Gaps 2;
40 AGAAGTGATCGCTGCCGTGGTGGCCATCGGGAAGTCGGATTTCTTACTCCCAAGGCTA 99
Db 32 AGAAGTGATCGCTGCCGTGGTGGCCATCGGGAAGTCGGATTTCTTACTCCCAAGGCTA 91
QY 100 TCGCCAAACAGGATCAAGTCCAAAGGGGCTGCGAAGCTACGGTGGTATGCGCAGATGTGCC 159
Db 92 TCGCCAAACAGGATCAAGTCCAAAGGGGCTGCGAAGCTACGGTGGTATGCGCAGATGTGCC 151
QY 160 AGAGCAGTGGCGGACAGAGATGGCTTTAGTGTCAATGTCATGTCGCAATCTCATCAGA 219
Db 152 AGAGCAGTGGCGGACAGAGATGGCTTTAGTGTCAATGTCATGTCGCAATCTCATCAGA 211
QY 220 GACAACTATTGCTGGCTTTCAGAAAATCCTCAGCAGTTTATGGATTATTTTTCAGAGGAAT 279
Db 212 GACAACTATTGCTGGCTTTCAGAAAATCCTCAGCAGTTTATGGATTATTTTTCAGAGGAAT 271
QY 280 TCCGAAATGACTTTCTAGAACTTCTCAGAGACGCTTTGGCACTAAAGGGTCCACAACA 339
Db 272 TCCGAAATGACTTTCTAGAACTTCTCAGAGACGCTTTGGCACTAAAGGGTCCACAACA 331
QY 340 ACATTGTCTACAAGAAATACATCAGCCACCGAGAGCACATCCACATGAATGCCATCAGT 399
Db 332 ACATTGTCTACAAGAAATACATCAGCCACCGAGAGCACATCCACATGAATGCCATCAGT 391
QY 400 GGGAACTCTGACTGATTTTACTAAGTGGCTGGGAGAGAGGGCTTGTGCAAGTGGACG 459
Db 392 GGGAACTCTGACTGATTTTACTAAGTGGCTGGGAGAGAGGGCTTGTGCAAGTGGACG 444
QY 460 AGACACCAAAAGGGCTGGTATATTCAGTACATAGACAGGGACCCAGAAACTATCCGCCGCG 519
Db 445 ----- 444
QY 520 AACTGGAACTGGAGAAAAGAAAAGCAGGACCTTGTATGATGAAGAAAAGAACTGCCAAAT 579
Db 445 ----- 444
QY 580 TTATTGAAGAGCAAGTGAGAGAGGCGCTGGAAGGAAGGAAACAGGAGGTCCCTACTTTTA 639
Db 445 ----- 444
QY 640 CGGAATTAAGCAGAGAAAATGATGAGAGAAAGTCACGTTTAAATTGAGTAAAGAGCAT 699
Db 445 ----- 444
QY 700 GTAGCTCATCGGAGCAACATCTTCCAAGTCAAGTACTCTGGACCGAGTGCACCTGAAGA 759
|||||

Db 445 -----AGTGCATGAAGA 457
QY 760 CGATAGGAAGTTTCAGCATCAGTGAACCGAAAAGAAATCTTCCCAGAGCTCAACTCAGTCTA 819
Db 458 CGATAGGAAGTTTCAGCATCAGTGAACCGAAAAGAAATCTTCCCAGAGCTCAACTCAGTCTA 517
QY 820 AAGAAAAGAGAAAAGAAAATCTGCACCTGGATGAAATCATGGAGATTTGAGAGGAAAAGA 879
Db 518 AAGAAAAGAGAAAAGAAAATCTGCACCTGGATGAAATCATGGAGATTTGAGAGGAAAAGA 577
QY 880 AAGAACTGCCCGCAACACAGCTACTTGGCTACAGCCTGAAATTTATTGTGAAAATTTATAACCA 939
Db 578 AAGAACTGCCCGCAACACAGCTACTTGGCTACAGCCTGAAATTTATTGTGAAAATTTATAACCA 637
QY 940 AAGAACTGGGAGAGAAAATATCATAGAAAAGGGCTATTGTTAAGGAAAGTAAATTGACAAAT 999
Db 638 AAGAACTGGGAGAGAAAATATCATAGAAAAGGGCTATTGTTAAGGAAAGTAAATTGACAAAT 697
QY 1000 ATACAGCTGTTGTAAGATGATTTGATTTCTGGAGACAAGCTGAAAATTTGACCGAGACTCAT 1059
Db 698 ATACAGCTGTTGTAAGATGATTTGATTTCTGGAGACAAGCTGAAAATTTGACCGAGACTCAT 757
QY 1060 TAGAGACAGTAATTTCCAGCACCGAGAAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAG 1119
Db 758 TAGAGACAGTAATTTCCAGCACCGAGAAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAG 817
QY 1120 GAAATGAAGGTACCTAGAAATCCATCAATGAGAAGACTTTTTCAGCTACTATCTCATTTG 1179
Db 818 GAAATGAAGGTACCTAGAAATCCATCAATGAGAAGACTTTTTCAGCTACTATCTCATTTG 877
QY 1180 AAACCTGCCCTTTTAAAGGAGCGCAGAGTTGAAAGAAATTCATATGAAGACATTTTCTAAAC 1239
Db 878 AAACCTGCCCTTTTAAAGGAGCGCAGAGTTGAAAGAAATTCATATGAAGACATTTTCTAAAC 937
QY 1240 TTGCTCAGTTTGAAAATTTGTTAAACAATA-CATTAAAAATCTTAAAGCATCAAAATTTGGTG 1298
Db 938 TTGCTCAGTTTGAAAATTTGTTAAACAATACCTTTAAATCTTAAAGCATCAAAATTTGGTG 997
QY 1299 TTCGC 1303
Db 998 TTCGC 1002
RESULT 6
ADQ56943
ID ADQ56943 standard; DNA; 679 BP.
XX AC ADQ56943;
XX DT 21-OCT-2004 (first entry)
XX DE Novel canine microarray-related DNA sequence SeqID8245.
XX KW canine microarray; drug screening; toxicity assay;
KW environmental pollutant; cellular response; gene expression profile;
KW toxic response; liver necrosis; fatty liver disease;
KW protein adduct formation; hepatitis; dog; ds.
XX OS Canis familiaris.
XX PN WO2004063324-A2.
XX PD 29-JUL-2004.
XX PF 05-MAY-2003; 2003WO-US013853.
XX PR 03-MAY-2002; 2002US-0377240P.
XX PA (GENE-) GENE LOGIC INC.
XX PA (PFIZ) PFIZER PROD INC.
XX PI Diggans JC, Porter M, Wei T;
XX

DR WPI: 2004-561890/54.
XX New isolated nucleic acid molecule, useful for drug screening and
PT toxicity assays or for assessing the impact, including toxicity, of a
PT compound, pharmaceutical agent or environmental pollutant on a cell or
PT living organism.
XX Claim 1; SEQ ID NO 8245; 41pp; English.
PS
XX This invention is related to a novel isolated canine nucleic acid
CC sequences and the construction of canine microarrays containing a
CC significant portion of the canine genome. The isolated canine nucleic
CC acid sequences of the invention may be useful for drug screening and
CC toxicity assays. The invention is therefore useful for assessing the
CC impact, including toxicity, of a compound, pharmaceutical agent or
CC environmental pollutant on a cell or living organism. The methods are
CC useful for detecting genes that are up- or down-regulated in canines in a
CC disease state. The sequences are useful as diagnostic agents or markers
CC to detect a cellular response in a sample individually or as part of a
CC gene expression profile. It is also useful as a target for agents that
CC modulate gene expression or activity. The database is useful for
CC producing electronic Northern blots that allow the user to determine the cell
CC type or tissue in which a given gene is expressed and to allow
CC determination of the abundance or expression level of a given gene in a
CC particular tissue or cell. The methods are useful for determining the
CC similarity of a toxic response to one or more individual compounds. The
CC methods are useful for predicting at least one toxic response or the
CC likelihood that a compound or test agent will induce various specific
CC pathologies such as those of the liver (liver necrosis, fatty liver
CC disease, protein duct formation or hepatitis), those of the kidney,
CC heart, brain or testes, or other pathologies associated with at least one
CC of the toxins. The methods are also useful for predicting or elucidating
CC the potential cellular pathways influenced, induced or modulated by the
CC compound or test agent due to the similarity of the expression profile
CC compared to the profile induced by a known toxin. The present sequence is
CC that of a canine DNA sequence which was claimed for use during the
CC production of a canine microarray of the invention.
XX
SQ Sequence 679 BP; 254 A; 106 C; 127 G; 185 T; 0 U; 7 Other;
Query Match 34.8%; Score 531.6; DB 13; Length 679;
Best Local Similarity 90.1%; Pred. No. 3e-118;
Matches 599; Conservative 0; Mismatches 61; Indels 5; Gaps 3;
QY 804 AGCTCAACTCAGTCTTAAGAAAAGAGAAAAGAAATCTGCAGTGAATCATGAGAG 863
DB 1 AGCTGGCTCAGTCAAAGAAAAGAAAAGAAAAGAAATCTGCNNCCGACGAAATCATNNAG 60
QY 864 ATTGAAGAGAAAAGAAAAGAACTGCCGACAGACTACTGGCTACAGCCTGAAATTATT 923
DB 61 ATTGAAGAGAAAAGAAAAGAACTGCCGACAGACTACTGGCTACAGCCTGAAATCATC 120
QY 924 GTGAAAATTATACCAAGAAACTGGGAGAGAAATATCATAAGAAAAGGCTATTGTTAAG 983
DB 121 GTGAAAATTATACCAAGAAACTGGGAGAGAAATATCATAAGAAAAGGCTATTGTTAAG 180
QY 984 GAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTTCTCGACAGCAAGCTGAAA 1043
DB 181 GAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTTCTCGACAGCAAGCTGAAA 240
QY 1044 CTTGACACAGACTCATTTAGACAGAGTAATTCAGACACGAGGAAAAGAAATCTAGTTTAA 1103
DB 241 CTTGACACAGACTCATTTAGACAGAGTAATTCAGACACGAGGAAAAGAAATCTAGTTTAA 300
QY 1104 AATTGGAGGCTACAGAGAAATGAAGTACCTAGATCCATCAATCAGAGAGCTTTTTC 1163
DB 301 AATTGGAGGCTACAGAGAAATGAAGTACCTAGATCCATCAATCAGAGAGCTTTTTC 360
QY 1164 GCTACTACTGCTATTGAAACTGGCCCTTTTAAAGGACGACAGTGTGAAGGAATTCATAT 1223
DB 361 GCTACTACTGCTATTGAAACTGGCCCTTTTAAAGGACGACAGTGTGAAGGAATTCATAT 420
QY 1224 GAAGACATTTCTAACTTGCTGCTGAGTTTGAATAATTTGTTTACATATTAATTAATCTTAA 1283

DB 421 GAAGACATTTCTAAACTTGCCTGAGTTTGAATACTTTTAAC---ACATTAATAATCTCTAA 477
QY 1284 AGCATCAAAATGGTGTTCGCCAAGGCATTATGAGACTCTACTGTGTAGGCTATATCTT 1343
DB 478 AACATCAAAATGGTGTTCGCCAAGGCATTATGAGACTCTACTGTGTAGGCTGT-TCCTT 536
QY 1344 TTGTATATAACCAACAGAGTTTTCGAAAATATTTACTGTATAGTTTGTTCAGCTAAACTTTGA 1403
DB 537 TTGTATATAAGCAATGGATTATTTAAATATTTACTGTATAGTTTGTTCAGCAAGCTTATA 596
QY 1404 GAAGAAATTAATATGCTCTCATGAGGTATCA-AACTATGTAATTTTGTCTTGTATTTT 1462
DB 597 GCAAAATCTAATATGCTCTCAGGAATGATCAGAACATGTAATTTTGTTCGTGTTTGT 656
QY 1463 TGTCTT 1467
DB 657 TCCTT 661
RESULT 7
AAH34095 ID AAH34095 standard; cDNA; 591 BP.
XX AC AAH34095;
XX 03-SEP-2001 (first entry)
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1177.
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX Homo sapiens.
XX WO200122920-A2.
XX 05-APR-2001.
XX 28-SEP-2000; 2000WO-US026524.
XX 29-SEP-1999; 99US-0157137P.
XX 03-NOV-1999; 99US-0163280P.
XX (HUMA-) HUMAN GENOME SCI INC.
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI: 2001-235357/24.
DR P-PSDB; AAG74690.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
PS Claim 1; Page 2977-2978; 9803pp; English.
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patient's own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027

```
CC to 1052, 7921 and 7922
XX Sequence 591 BP; 199 A; 95 C; 114 G; 182 T; 0 U; 1 Other;
SQ Best Local Similarity 32.1%; Score 490.8; DB 4; Length 591;
Matches 512; Conservative 0; Mismatches 12; Indels 4; Gaps 1;

QY 1005 GCTGTTGGAAGATGATTCCTGGAGCAAGCTGGAACCTGACCAAGCTCATTTAGAG 1064
Db 28 GCTGAGGAATTCGGCAGGCTCTGGAGCAAGCTGGAACCTGACCAAGCTCATTTAGAG 87
QY 1065 ACAGTAATTCAGACACAGGAAAGAAATTCAGTTTAAATGGAGGCTACAGAGGAAT 1124
Db 88 ACAGTAATTCAGACACAGGAAAGAAATTCAGTTTAAATGGAGGCTACAGAGGAAT 147
QY 1125 GAAGGTACCTTAGAATCCATCAATGAGAAGACTTTTTCAGCTACTATCGTCATTGAAACT 1184
Db 148 GAAGGTACCTTAGAATCCATCAATGAGAAGACTTTTTCAGCTACTATCGTCATTGAAACT 207
QY 1185 GGCCCTTTAAAGGACGAGAGTTGAAGGAATTCATATGAGACATTTCTAAACTTGCC 1244
Db 208 GGCCCTTTAAAGGACGAGAGTTGAAGGAATTCATATGAGACATTTCTAAACTTGCC 267
QY 1245 TGAGTTGAAAATTTGTTAAACAATACATTAATACTTAAAGCATCAAAATTTGGTTCGCC 1304
Db 268 TGAGTTGAAAATTTGTTAAACAATACATTAATACTTAAAGCATCAAAATTTGGTTCGCC 327
QY 1305 AAGCATTTATGAGACTCTACTGTGTAGGGTATATCTTTTGTATATAAACAAGGTTT 1364
Db 328 AAGCATTTATGAGACTCTACTGTGTAGGGTATATCTTTTGTATATAAACAAGGTTT 387
QY 1365 TTGAAAATATTACTGTA----TAGTTGTTGAGCTAAACTTTTGAGAAGAAATTTAATTATGT 1420
Db 388 TTGAAAATATTACTGTAATAGTTAGTTGTTGAGCTAAACTTTTGAGAAGAAATTTAATTATGT 447
QY 1421 CTCATGAGGTATCAAACTATGTAATTTGTCCTGTTATTTTGTCTTTGTTGTAATTTA 1480
Db 448 CTCATGAGGTATCAAACTATGTAATTTGTCCTGTTATTTTGTCTTTGTTGTAATTTA 507
QY 1481 CTTGATGAGTTATATCTTCAATTAAGAATGTTATTAATAAAAAA 1528
Db 508 CTTGATGAGTTATATCTTCAATTAAGAATGTTATTAATAAAAAA 555

RESULT 8
ID AAX85550
AX AAX85550 standard; cDNA; 1102 BP.
AC AAX85550;
XX
XX 07-SEP-1999 (first entry)
DT
DE cDNA of a gene coding for a mouse deleted kin17 protein.
XX
KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW protein interaction; curved DNA; HIV replication; HIV integration;
KW repair enzyme; ss.
XX
OS Mus sp.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX
```

```
DR WPI; 1999-359999/31.
XX New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.
XX Claim 4; Page 31; 69pp; French.
XX
CC The present sequence encodes a mouse kin17 protein with amino acids 129-
CC 228 deleted. The mammalian kin17 protein is useful for preparing a
CC medicament for controlling cell proliferation or for controlling
CC fertility. The medicaments can also be used to treat hyperproliferative
CC diseases. Fragments between amino acids 55 and 235 (preferably between
CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
CC regulating the interaction between proteins and curved DNA. The fragment
CC can be used to block replication of HIV or its integration into the human
CC genome or to target repair enzymes to curved DNA sites. Expression
CC vectors for kin17 can be used for controlling cell proliferation
XX
SQ Sequence 1102 BP; 373 A; 205 C; 272 G; 252 T; 0 U; 0 Other;

Query Match 31.1%; Score 475; DB 2; Length 1102;
Best Local Similarity 65.4%; Pred. No. 1.8e-104;
Matches 897; Conservative 0; Mismatches 180; Indels 294; Gaps 2;

QY 66 ATGGGGGAAGTCGGATTTTCTTACTCCCAAGCTATCCCAACAGATCAAGTCCAAGGG 125
Db 1 ATGGGCAAGTCGGATTTTCTGAGCCCAAGGCCATCGCCAATGAATTAAGTCCAAGGG 60
QY 126 CTGCAGAAAGCTACGCTGGTATTGGCAGATGTCCAGAAAGCAGTCCCGGACGAGAATGGC 185
Db 61 CTCAGAAAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCCGACGAGAATGGC 120
QY 186 TTTAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAACTATTGTGCTGGCTTCAGAAAT 245
Db 121 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAGACAACTGTGTGCTGGCTTCAGAAAC 180
QY 246 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTC 305
Db 181 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTTCCGAATGACTTTCTGGAACCTTCTG 240
QY 306 AGGAGACGCTTTTGGCACTTAAAGGGTCCCAACAACTTGTCTACAAAGATACATCAGC 365
Db 241 AGGCGACGCTTTGGCACTTAAAGGGTCCCAACAACTTGTCTACAAATGAATACATCAGC 300
QY 366 CACCGAGAGCACATCCACATGAATGCCATCAGTGGGAAAACCTCTGACTGATTTTACTAAG 425
Db 301 CACCGAGAGCACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAG 360
QY 426 TGGCTGGGACAGAAAGGCTTGTGCANAAGTGGACGAGACACCAAAAGGCTGGTATATTTCAG 485
Db 361 TGGCTGGGACAGAAAGGCTTGTGTAAA-----TGTTGAAA----- 387
QY 486 TACATAGACAGGGACCCAGAAAACCTATCCGCGGCAACTGGAACTGGAGAAAAAGAAAAG 545
Db 388 ----- 387
QY 546 CAGGACCTTGATGATGAAGAAAAAAACCTGCCAAATTTATTGAAGAGCAAGTGAGAAGAGGC 605
Db 388 ----- 387
QY 606 CTGGAAGGGAAGGAACAGGAGGTCCTTACTTTTACGGAATTAAGCAGAGAAAATGATGAA 665
Db 388 ----- 387
QY 666 GAGAAAAGTCACGTTTAAATTTTGAAGTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCC 725
Db 388 ----- 387
QY 726 AAGTCAAGTACTCTGGGACCGAGTGCATCTGAAGACGATAGGAAGTTCAGCATCAGTGAAG 785
Db 388 -----GGTACAGTGCATCTGAAGCTGCTGGGGAGCGCAGCAGTCCGGAAA 432
QY 786 CGAAAAGAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCA 845
```

Db 433 CGGAAGAGCTCTCAGAGCTCCGCCAGCT-----CGGAAGAGAGAGTCCGCC 486
 Qy 846 CTGGATGAATCATCGAGATTGAAGGAGAAAGAAAGAACTGCCCGACAGACTACTGG 905
 Db 487 CTGGATGAGATCATCGAGCTCGAAGAGGAGAAAGAAAGAGCCGACGACAGCCCTGG 546
 Qy 906 CTACAGCTGAATATTCTGTAATTAATTAACCAAGAACTGGGAGAGAAATATCATAG 965
 Db 547 TTACAGCCGGGATCGTTGTGAATTTAATGAAGAGCTTTGGGAGAGAAATATCACAG 606
 Qy 966 AAAAGGCTATTGTTAAGGAAGTAAATGAACAATATACAGCTGTTGTGAAGATGATTAT 1025
 Db 607 AGAAGGGGTGTTAAGGAAGTAAATGAACAAGTACAGAGTACAGAGTGTGTTAAGATGACTGAC 666
 Qy 1026 TCTGGAGACAACTGAACTGACAGACTCAATTTAGACAGATTAATTTCCAGCAGCAGGA 1085
 Db 667 TCTGGAGACAGGCTGAACTGGACCAAGACTCATTTAGACAGACTCAATTTCCGCCCGCGGG 726
 Qy 1086 AAAAGAACTTCTAGTTTAAATGGAGCTACAGAGAAATGAAGTACCTTAGAATCCATC 1145
 Db 727 AAAAGGGTCTAGTTTAAATGGAGCTACAGAGAAATGAAGGCACTCTCGAATCCATC 786
 Qy 1146 AATGAGAGACTTTTTCAGCTACTATCGTCAATTTGAACTGGCCCTTTAAAGGAGCGCAGA 1205
 Db 787 AATGAGAGGCTTTTTCAGCCAGATAGTCAATTTGAACTGGACCTTTGAAAGGAGCGCAGA 846
 Qy 1206 GTTGAAGGAATCAATATGAAGACATTTCTAACTTGCCTGAGTTTGAAGAAATTTGTTAAC 1265
 Db 847 GTTGAAGGTATCAATATGAAGACATCTAACTTGCCTGAGTTTGAAGAAATTTGTTAAC 906
 Qy 1266 AATACATTAATTTAAAGCATCAATTTGGTGTTCGCCAAGGCATTTAGACTCTACT 1325
 Db 907 AACACATTTGAACTGTGAAGCATCAATTTGGTGTTCGCCAAGGCATTTAGACTCTACT 966
 Qy 1326 GTGTTAGGCTATTTCTTTGTATATAAACAAGAGGTTTTTGAAGAAATATTTAGTATAGT 1385
 Db 967 GTGTTAGGGAATTTGTTGTATATAAAGAAAGAAATATTTAATTAATTAAGTATAGT 1026
 Qy 1386 TGTTCAGCTAACTTTGAGAAGAAATTTAATTTATGTTCTCATGAGGTATCAAA 1436
 Db 1027 TAGTTCGGTAAATTTATATAAATATCTATGTTTGTATTAAGTGAAGAAAT 1077

RESULT 9

ADP22451

ID ADP22451 standard; DNA; 1372 BP.

XX AC ADP22451;

XX 12-AUG-2004 (first entry)

XX DE Sea-squirt (Ciona intestinalis) zinc finger protein coding sequence #13.

XX DE sea-squirt; zinc finger protein; gene detection; drug development;

XX KW zinc finger protein-associated disease; gene; ds.

XX OS Ciona intestinalis.

XX XX JP2004057126-A.

XX PD 26-FEB-2004.

XX PF 31-JUL-2002; 2002JP-00222484.

XX XX 31-JUL-2002; 2002JP-00222484.

XX PR (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX PA WPI; 2004-208711/20.

XX DR P-PSDB; ADP22452.

XX XX Novel gene encoding zinc finger protein, useful as probe in gene

PT

PT detecting instruments and in development of drug for treating zinc finger
 PT protein associated diseases.

PS Claim 3; SEQ ID NO 25; 972pp; Japanese.

CC The invention comprises the amino acid and coding sequences of sea-squirt
 CC (Ciona intestinalis) zinc finger proteins. The DNA and protein sequences
 CC of the invention are useful in a gene detecting instrument. The DNA and
 CC protein sequences of the invention are useful in the development of drugs
 CC for the treatment of zinc finger protein-associated diseases. The present
 CC DNA sequence encodes a sea-squirt zinc finger protein of the invention.

SX Sequence 1372 BP; 500 A; 224 C; 286 G; 362 T; 0 U; 0 Other;

Query Match 24.9%; Score 380.6; DB 12; Length 1372;

Best Local Similarity 60.0%; Pred No. 1.3e-81;

Matches 704; Conservative 0; Mismatches 439; Indels 30; Gaps 3;

Qy 69 GGGAAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAAGGATCAAGTCCAAAGGGCTG 128

Db 52 GAGNAACCAAGGATTTCTCACTGCCAAGGCTATCGGCAACCGTATCAAAATCAAAAGGACTT 111

Qy 129 CAGAAAGCTACGCTGTGATTTGCCAGATGTGCCAAGAGCAGTCCCGGACGAGATTCGCTTT 188

Db 112 CAAAAGTTCCGATGTTGTTCAATGTGCCAAAACAGTGCAGAGATGAGAACCGGTTT 171

Qy 189 AAGTCTCATTTGATGTCGGAATCTCATCAGACAACTATTGCTGCTTCAGAAAATCTCTCAGG 308

Db 232 GGAGCTTTTACAGTTTCTTTTCAAGTTCCTTTTCAAGACTTTTATGCAACTTTTGAAA 291

Qy 309 AGACGCTTTGGCACTTAAAGGGTCCACAAACATTTGTCTACAAAGATATACATCAGCCAC 368

Db 292 ACACGCTTTCCGAGCTTAAAGAGTTTCAACAATTTGTGTATACAGATATATTTCCCAT 351

Qy 369 CGAGAGCATCCACATGAATGCCACTCAGTCGGAAACTCTCAGCTGATTTTACTAAGTGG 428

Db 352 AAAGAACACATTCATATGAACCGTCAAAATGGGTGACGTTAACTGCTTTCACAAATGG 411

Qy 429 CTGGGCGAGAAAGGCTTTGTGCAAGTGGACGAGACACCAAAAGGCTGTATTTCAAGTAC 488

Db 412 TTGGGTAGAGAGGACACTGTAAAGTGGATCAACCGAAAGGGTTGGTTTCATACATAC 471

Qy 489 ATAGACAGGACCCAGAAACTATCCCGCGCAACTGGAACCTCGAGAAAAGAAAGACAG 548

Db 472 ATTGATAGAGACCCATTTGGTGTGCGCCACAGAAAGAACTTTGATAAACAAGCAAGG 531

Qy 549 GACCTTGATGATGAGAAAGAAACTGCCAAATTTTATTGAAGAGCAAGTGAAGAGAGCCCTG 608

Db 532 GAGAGAGATGATGATGAAGAGCAAGCAAAAGCAATTTGAAGAAATGTTTAAAGAGACAA 591

Qy 609 GAAGGAGGAAGAAACAGGAGGTCCTTACTTTTACGGAATTAAGCAGAGAAATGATGAAGAG 668

Db 592 GA-----TACTGCTAAGGATTCAGACTCCCGGAAAAATTTACTGAG 630

Qy 669 AAAGTCACTGTTTAAATTTGATGAAGAGAGATGTAGCTCATCCGAGCAACATCTTCCAAG 728

Db 631 TTGGTTCCGAATGATGAGGAAAAAATTCGAATAAATCT-----GCAAAATGAATTCGAG 684

Qy 729 TCAAGTACTCTGGGACCGAGTGCACCTGAAGAGATAGGAAGTTCACATCAGTGAAGACGA 788

Db 685 CCAAAAGCTGCCACATCTTTCACAGCTTGCACCTTCGTGTGCTCAGCTTATCCAGTTTGGTGGC 744

Qy 789 AAAGAATCTTCCAGAGCTCAACTCAGTCTTAAGAAAGAAAGAAAGAAATCTGCACCTG 848

Db 745 AAGAGAAACAGCAAGCTCATCCGCTTTTAAACCTAGCAAGAGATCAAAAGTCTGCACCTG 804

Qy 849 GATGAATCATGGAGATGAAGAGG---AAAAGAAAGAACTGCCCGAAGACAGACTACTGG 905

Db 805 GATGAATCATTTGACGCTGAGGAGGAAAGAAAGAGCAGAGTTTACACACTGAACACTGG 864

PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Claim 1; SEQ ID NO 39070; 2lpp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins (AB157737-
CC AB172072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1241 BP; 371 A; 294 C; 352 G; 224 T; 0 U; 0 Other;
Query Match 19.5%; Score 298.6; DB 4; Length 1241;
Best Local Similarity 54.2%; Pred. No. 8.e-62;
Matches 638; Conservative 0; Mismatches 524; Indels 15; Gaps 1;
64 CCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAAGG 123
67 CGATGGGTCGCGCGAGGTAGGTAGCGCCCAAGTACTCGCCCAACAGATGAATCGAAGG 126
124 GCCTCAGAGCTACGCTGGTATTGCCAGATGTCGAGAGCAGTCCGCGGACGAGAATG 183
127 GTCTCAGAAGTCGCTGGTACTGCCAGATGTGCGAAGCAGTCCCGCGATGAAGACG 186
184 GCTTTAAGTGTCAATGTATGTCGGAATCTCATCAGACAACTATTGTGCTGGCTTCAGAAA 243
187 GCTTCAAGTGCACAGGATGAGGAGTCCACAGCGCCAGTGTCTCTTTGGGACA 246
244 ATCTCTCAGCAGTTTATGGATTATTTTTCAGAGAAATTCGGAATGACTTTCTAGAACTTC 303
247 ATCTCTGGCAATTCCTGCACAGCTTCAGCAAGAGTTCCTCCGACGCTACATGGAGTTGC 306
304 TCAGGAGACGCTTTGGCATTAAAGGTTCCACAAACAACTTCTTACACGATATCATCA 363
307 TCGCGCGCGGTTTCGACAGCAAGCGCAACAGCGCCCAACAGATCTACAGGAGTACATG 366
364 GCCACCGAGAGCAGATCCACATGATGTCACATGTCAGTGGGAACTCTGACTGATTTTACTA 423
367 CCACAAAGGAGACATCCACATGAAACGCCACCGATGGCTCACCCTGTCGACTACGTGA 426
424 AGTGGCTGGGCGAGAGAGGCTTTGTGCAAGTGGACGAGACACAAAGGCTGGTATATTC 483
427 AGTGGCTGGGCGGACTGGGCAAGTCATAGCGGATGAGACGGAAGGCTGGTTCGTCA 486
484 AGTACATAGACAGGAGCCAGAAAATCTATCGCGCGCACTGGAACTGGAGAAAAGAAAA 543
487 CCTACATTGTCAGTCCAGAGGCCATGGAACGACAGCGCAAGGCTGATCGCAAGGAGA 546
544 AGCAGGACCTTGTATGATGAAGAAAATCTGCCAAATTTATTGAAGAGCAAGTGAAGAG 603
547 AGATGGAAGAGGACGAGGAGCGGATGGCCGACTTCATTGAGCAGCAGATTAATAATG 606
604 GCCTGGAAGGGAAGGAGGAGGCTCCCTACTTTTACGGAATTAAGCAGAGAAAATGATG 663
607 CCAAGCCRAAGGACGCGGAGGAGGACGAGCCAGGAGAAGTTCACCGAGCTAAAGCGCG 666
664 AAGAGAAAGTCAGTTTAATTTAGTAAAGGACATGTAGCTCATTCGGGAGCAACATCTT 723
667 AAGAGAAACGCAACCACTCAAGCTTGTATATTCG-----CCTTGAGAAAAGT 711
724 CCAAGTCAAGTACTCTTGGGACCGAGTGCACCTGAAGACGATAGGAAGTTCAGCATCATGTA 783

Db 712 TCCAGCCTGACACTGTGCTAGGGAAATCCGCTCTAGCCAAAGCAGCTGCCCTCGAAGCTG 771
Qy 784 AACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAGAAAAGAAATCTG 843
Db 772 AAGAAAGGTGTTCAAGAAACCCAAATCCGTGGTGGAGACAGCAAAACGCGGTCCGTGC 831
Qy 844 CACTGGATGAATCATGGAGATTGAAGAGGAAAAGAAAAGAACTGCCCGGAACAGACTACT 903
Db 832 TGGACGAGATCATCAAGCAGGAGGAAAAGGAAAGGAGCGTGCACAAACCGCAAGGACTACT 891
Qy 904 GGCTACAGCCCTGAATTTATTTGTAAGAAATTAACCAAGAACTGGGAGAGAAATATCATA 963
Db 892 GGCTGCACAAGGGTATCGTGGTCAATTTATTTCCAAATCCATGGGCGAAAAGTTCTTCA 951
Qy 964 AGAAAAGGCTATTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGTAAGATGATTG 1023
Db 952 AACAAAAGCGGTTGCTCGACGATTAATGACAGATATCAGGSCAAAATCAAGTTCTTGG 1011
Qy 1024 ATCTGGAGACAAGCTGAAACTTCACCAGACTCATTTTAGACACAGTAAATTCAGCACCAG 1083
Db 1012 AGACTGGGGAAGCTAAAGTGGATCAAGCTCATTTTGGAGACGCTAATCCCGCTTTGG 1071
Qy 1084 GAAAAGAAATCTAGTTTAAATGGAGCTACAGAGGAAATGAAGGTACCTAGATCA 1143
Db 1072 ACAAGCCTGTGCTGGTGGTTAATGGCGCTTATCGGGGATCCGAGGCTCTGCTAAGGAAAC 1131
Qy 1144 TCAATGAGAAGACTTTTTCAGCTACTATCGTCAATTGAAACTGGGCCCTTTAAAAGGACGCA 1203
Db 1132 TGGACGCGGAGATATTCAGTACGCTGGGAATATTTGCACGCTCTCTCAAGGCGAGAA 1191
Qy 1204 GAGTTGAAGGAATTCATATATGAAGACATTTCTAAACT 1240
Db 1192 TTGTAGACAACGTCGCAATACGAAGATATATCTAAACT 1228
RESULT 12
AAAA1559
ID AAA01559 standard; cDNA; 300 BP.
XX AC AAA01559;
XX DT 19-MAY-2000 (first entry)
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:1550.
XX Human; colon cancer; tumour; diagnosis; gene expression product; probe;
KW detection; cancerous state; metastasis; identification; breast cancer;
KW oestrogen receptor-positive breast cancer; therapy;
XX oestrogen receptor-negative breast cancer; lung cancer; ss.
OS Homo sapiens.
XX WO9958675-A2.
XX PD 18-NOV-1999.
XX PF 13-MAY-1999; 99WO-US010602.
XX PR 14-MAY-1998; 98US-0085426P.
XX PR 15-MAY-1998; 98US-0085537P.
XX PR 15-MAY-1998; 98US-0085696P.
XX PR 21-OCT-1998; 98US-0105234P.
XX PR 27-OCT-1998; 98US-0105877P.
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Lebat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI; 2000-126369/11.

PR	26-JUL-1999;	99US-0145276P.
PR	27-JUL-1999;	99US-0145913P.
PR	27-JUL-1999;	99US-0145918P.
PR	27-JUL-1999;	99US-0145919P.
PR	28-JUL-1999;	99US-0145951P.
PR	02-AUG-1999;	99US-0146386P.
PR	02-AUG-1999;	99US-0146388P.
PR	02-AUG-1999;	99US-0146389P.
PR	03-AUG-1999;	99US-0147038P.
PR	04-AUG-1999;	99US-0147204P.
PR	04-AUG-1999;	99US-0147302P.
PR	05-AUG-1999;	99US-0147192P.
PR	05-AUG-1999;	99US-0147260P.
PR	06-AUG-1999;	99US-0147303P.
PR	06-AUG-1999;	99US-0147416P.
PR	09-AUG-1999;	99US-0147933P.
PR	10-AUG-1999;	99US-0148171P.
PR	11-AUG-1999;	99US-0148319P.
PR	12-AUG-1999;	99US-0148341P.
PR	13-AUG-1999;	99US-0148565P.
PR	13-AUG-1999;	99US-0148684P.
PR	16-AUG-1999;	99US-0149368P.
PR	17-AUG-1999;	99US-0149175P.
PR	18-AUG-1999;	99US-0149426P.
PR	20-AUG-1999;	99US-0149722P.
PR	20-AUG-1999;	99US-0149723P.
PR	20-AUG-1999;	99US-0149929P.
PR	23-AUG-1999;	99US-0149902P.
PR	23-AUG-1999;	99US-0149930P.
PR	25-AUG-1999;	99US-0150566P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-0152363P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155659P.
PR	28-SEP-1999;	99US-0156458P.
PR	29-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-0157753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0159293P.
PR	13-OCT-1999;	99US-0159294P.
PR	13-OCT-1999;	99US-0159295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159584P.
PR	21-OCT-1999;	99US-0160741P.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	25-OCT-1999;	99US-0161404P.
PR	25-OCT-1999;	99US-0161405P.
PR	25-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	26-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Query Match 14.7%; Score 224.4; DB 3; Length 1459;		
Best Local Similarity 53.6%; Pred. No. 8.4e-44;		
Matches 662; Conservative 0; Mismatches 516; Indels 58; Gaps 7;		
Qy	66	ATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCCCAACAGGATCAAGTCCAAGGGG 125
Db	1	ATGGGTAAAGAACGATTTTCTAAACCCCAAGCGATTCGAATCGAATTAAGCCCAAGGA 60
Qy	126	CTGCAGAAAGCTACGCTGGTATTGGCCAGATGTCCAGAAAGCAGTCCCGGACGAGAAATGGC 185
Db	61	CTTCAAAAGCTTCGATCGTATTGTTCAGATGTGTAGAAACAATGCCAGAGCGAGAATGGA 120
Qy	186	TTTAAAGTGTCAATTGTATGTCCGAATCTCATCAGAGACAACTATTGTCTGGCTTCAGAAAT 245
Db	121	TTCAAGTGTCAATTGTATGAGCGAATCTCACAGAGGCAGATGCNAAGTTTGTGGCAGAT 180
Qy	246	CCTCAGCAGTTTATGGATTAATTTTTCAGAGAAATTCGCAAAATGACTTTCTAGAACTTCTC 305
Db	181	CCGACTCGTGTGTCGATGGTACTCTCTGAAGAAATTCGAGCAGACGTTTCTCGATCTGATG 240
Qy	306	AGGAGACGCTTTGGCACTAAAGGTTCCACAACAACTTGTCTACAAGATACATCAGC 365
Db	241	AGCGGAGCCACCGCTTTCTCTCGTATCGCTGCCACTGTTGTCTACAATGAGTATATTAAAC 300
Qy	366	CACCGAGACACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTTAAG 425
Db	301	GATAGACATCATGTTTCACATGAATTCGACGGAGTGGGCGAGCTTGACGGAGTTTATCAAG 360
Qy	426	TGGCTGGGACAGAAAGCTTTGTGCAAAAGTGGACGAGACACCAAAAGCTGGTATATTCAAG 485
Db	361	CATCTTGGGAAACTCGTAAAGTGTAAAGTTGAAGAAACTCCTAAAGGTTGGTTTTATTACT 420
Qy	486	TACATACAGAGGACCCAGAAACTATCCGCGGCACTGGAACCTGGAGAGAAAAAGAAAG 545
Db	421	TATATAGATAGAGATTCAGACTTTGTTTAAAGGAGAGGTTGAAGAAATAGAGAGTTTAAG 480
Qy	546	CAGGACCTTGATGATGAAGAAAAAACTGCCAAATTTATTGA-----AGAGCA 592
Db	481	AGTGATTTAGCTGAGGAGGAGAGCAAGAGGGAGATTCAGAGACAGATTGAGAGAGCT 540
Qy	593	AGTGAGAAAGGCGCTGGAAGGGAAGNAACAGGAGGTCCTTACTTTTACG----- 641
Db	541	GCTGAGAAATTGAATGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 600
Qy	642	-----GAATTAACGAGAGAAATGATGAAGAG-AAAAGTCACG 677
Db	601	GATGATGGTGATGATGAGAGGAAAGAAAGACTTGAAGTTTGAAGAGTGGGGTCAAG 660
Qy	678	TTTAAATTTAGTAAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCNAGTCAAGT--A 735
Db	661	GTTGGATTTGCTCTTGTGTGGAGGTGTTAAACAGGTTGCTACGGGTAAGAGAGAGGGGAA 720
Qy	736	CTCTGGGACCGAGTGCACCTGAAGACGATAGGAAGTTTCAGCATCAGTGAACGAAAGAAAT 795
Db	721	AGCTCGAAACTTTTGTGTTGGGATGAAGAGAAATGACAAAGTTGAGAGAGGGGAGAGAGG 780
Qy	796	CTTCCCAGAGCTCACTCAGTCTTAAGAAAGAGAAAGAAAGAAATCTGCATGGATGAA 855
Db	781	AAGAGAAAGTGGGAGCTCCGG--GAGGAGCGAGAGGAGAGAGATCAGCTTTTGGATGAGT 838
Qy	856	TCATGGAGATTGAAGAGGAAAAAGA---AAAGAACTGCCCGAAACAGACTACTTGGCTACAGC 912

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:58:53 ; Search time 8541.29 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-1

Perfect score: 1528

Sequence: 1 ctgaattcagggcgctg.....atgtattataaaaaaa 1528

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_hc: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_est7: *
9: gb_gsa1: *
10: gb_gsa2: *
11: gb_gsa3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1396	91.4	1396	4	CR595908 full-length
2	1199.4	78.5	1339	4	CR618602 full-length
3	1189.2	77.8	1543	4	AY609991 Sus scrofa
4	1065	69.7	1182	10	AY412518 Homo sapi
5	1039.8	68.0	1182	10	AY412519 Pan trogl
6	900	58.9	965	7	AL522388 AL522388
7	859.2	56.2	951	7	CO725903 ILLUMIGEN
8	849	55.6	988	7	CO726806 ILLUMIGEN
9	831	54.4	846	1	AL558811 AL558811
10	800.4	52.4	882	3	BQ221694 AGENCOURT
11	767.8	50.2	1154	10	AY412520 Mus muscu
12	757.4	49.6	1944	1	AL522387 AL522387
13	744.6	48.7	842	3	BI862461 603391378
14	744.2	48.7	936	6	CD725073 ILLUMIGEN
15	742.6	48.6	890	6	CD109094 AGENCOURT
16	741	48.5	768	7	CR986081 CR986081
17	735.8	48.2	848	1	AL558810 AL558810
18	728	47.6	748	5	BU161074 AGENCOURT
19	720.2	47.1	922	6	CD389584 AGENCOURT
20	715.4	46.8	718	8	CX866441 HBSC4_10
21	709.4	46.4	733	5	BX104466 BX104466
22	686.8	44.9	723	3	BQ186497 UI-E-EJ1-

23	683	44.7	693	7	CN256425	CN256425	170004245
24	678.6	44.4	1044	7	CO724635	CO724635	ILLUMIGEN
25	677.8	44.4	710	1	AM089573	AM089573	xd20e01.x
26	671.8	44.0	720	7	CK301009	UI-E-EJ1-	CK301009
27	651.8	42.7	666	1	AL700007	DKE20686M	AL700007
28	647	42.3	783	7	CN166221	997480 MA	CN166221
29	646.8	42.3	656	7	CV029155	7852 Full	CV029155
30	642.8	42.1	786	1	AJ819816	AJ819816	AJ819816
31	642.4	42.0	701	1	AI807250	wf38f09.x	AI807250
32	637	41.7	864	1	AM037867	AM037867	AM037867
33	635.2	41.6	790	8	DN523891	1267468 M	DN523891
34	632.4	41.4	666	3	BM739997	K-EST0010	BM739997
35	630.4	41.3	1011	7	CO725532	ILLUMIGEN	CO725532
36	630.2	41.2	871	1	AJ819723	AJ819723	AJ819723
37	627.6	41.1	682	1	AV721396	AV721396	AV721396
38	627.4	41.1	638	3	BM711163	UI-E-DX1-	BM711163
39	622	40.7	634	3	BM738984	K-EST0008	BM738984
40	612.8	40.1	698	1	AI650375	wa90b01.x	AI650375
41	608.4	39.8	942	5	BQ714508	AGENCOURT	BQ714508
42	602	39.4	719	7	CN792844	4127778 B	CN792844
43	585.2	38.3	1077	5	BU511663	AGENCOURT	BU511663
44	576.4	37.7	695	8	DN526620	1271404 M	DN526620
45	576.2	37.7	605	6	CB242699	UI-CF-PNO	CB242699

ALIGNMENTS

RESULT 1	CR595908	1396 bp	mRNA	linear	HTC 21-JUL-2004
LOCUS	full-length cDNA clone CS0DJ015Y119 of T cells (Jurkat cell line)				
DEFINITION	Cot 10-normalized of Homo sapiens (human).				
ACCESSION	CR595908.1	GI:50476715			
VERSION	HTC; CNSLT CDNA.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1396)				
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished				
REMARK	Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ invitroGen Corporation 1600 Paraday Avenue				
REFERENCE	2 (bases 1 to 1396)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.				
FEATURES	Location/Qualifiers				
source	1..1396				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="CS0DJ015Y119"				
	/tissue_type="T cells (Jurkat cell line) Cot 10-normalized"				
	/plasmid="pCMVSPORT_6"				
ORIGIN					
Query Match	91.4%	Score 1396;	DB 4;	Length 1396;	
Best Local Similarity	100.0%;	Pred. No. 6.2e-309;	Indels 0;	Gaps 0;	
Matches 1396;	Conservative 0;	Mismatches 0;			

```
Qy 67 TGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAAAGGGCC 126
Db 1 TGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAAAGGGCC 60
Qy 127 TGCAGAGCTACGCTGGTATTTGCCAGATGTCGCCAGACAGTCCGGGACGAGATGGCT 186
Db 61 TGCAGAGCTACGCTGGTATTTGCCAGATGTCGCCAGACAGTCCGGGACGAGATGGCT 120
Qy 187 TTAAGTGTCAATGTATGTCGGAATCTCATCAGAGCAACTATTGCTGGCTTCAGAAAAATC 246
Db 121 TTAAGTGTCAATGTATGTCGGAATCTCATCAGAGCAACTATTGCTGGCTTCAGAAAAATC 180
Qy 247 CTCAGCAGTTTATGGAATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTCA 306
Db 181 CTCAGCAGTTTATGGAATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTCA 240
Qy 307 GGAGACGCTTTGGCACTAAAGGGTCCACACACATTTGCTACACGAATACATCAGCC 366
Db 241 GGAGACGCTTTGGCACTAAAGGGTCCACACACATTTGCTACACGAATACATCAGCC 300
Qy 367 ACCGAGAGCACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGT 426
Db 301 ACCGAGAGCACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGT 360
Qy 427 GCCTGGGCAGAGAAGGCTTGTGCAAAAGTGGACGAGACACCAAAAGGCTGGTATATTCAGT 486
Db 361 GCCTGGGCAGAGAAGGCTTGTGCAAAAGTGGACGAGACACCAAAAGGCTGGTATATTCAGT 420
Qy 487 ACATAGACAGGACCCAGAACTATCCGCCGCACTCGAACTGGAGAAAAGAAAAGC 546
Db 421 ACATAGACAGGACCCAGAACTATCCGCCGCACTCGAACTGGAGAAAAGAAAAGC 480
Qy 547 AGGACCTTGATGATGAAGAAAAAATCGCCAAATTTATTGAAGAGCAAGTGAAGAGAGGCC 606
Db 481 AGGACCTTGATGATGAAGAAAAAATCGCCAAATTTATTGAAGAGCAAGTGAAGAGAGGCC 540
Qy 607 TGGGAAGGAAGAACAGAGAGTCCCTACTTTTACGGAATTAAGCAGAGAAAATGATGAAG 666
Db 541 TGGGAAGGAAGAACAGAGAGTCCCTACTTTTACGGAATTAAGCAGAGAAAATGATGAAG 600
Qy 667 AGAAGTCACGTTTAAATTTAGTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCCA 726
Db 601 AGAAGTCACGTTTAAATTTAGTAAAGGAGCATGTAGCTCATCCGGAGCAACATCTTCCA 660
Qy 727 AGTCAAGTACTCTGGGACCGAGTGCATGAAGAGCATAGGAAGTTTCAGCATCAGTGAAC 786
Db 661 AGTCAAGTACTCTGGGACCGAGTGCATGAAGAGCATAGGAAGTTTCAGCATCAGTGAAC 720
Qy 787 GAAAAGAACTCTCCAGAGCTCAACTCAGTCTAAAGAAAAGAAAAGAAATCTGCAC 846
Db 721 GAAAAGAACTCTCCAGAGCTCAACTCAGTCTAAAGAAAAGAAAAGAAATCTGCAC 780
Qy 847 TGGATGAAATCATGAGATTCAAGAGGAAAAGAAAGAACTGCCCGAACACACTACTGGC 906
Db 781 TGGATGAAATCATGAGATTCAAGAGGAAAAGAAAGAACTGCCCGAACACACTACTGGC 840
Qy 907 TACAGCCTGAAATTTATGTGAAAAATTATAACCAAGAACTGGGAGAGAAAATATCATAAGA 966
Db 841 TACAGCCTGAAATTTATGTGAAAAATTATAACCAAGAACTGGGAGAGAAAATATCATAAGA 900
Qy 967 AAAAGGCTATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGGAAGATGATGATT 1026
Db 901 AAAAGGCTATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGGAAGATGATGATT 960
Qy 1027 CTGAGAGCAAGCTGAACTCCACAGACTCATTTAGACACAGTAATTCACGACACAGAA 1086
Db 961 CTGAGAGCAAGCTGAACTTGACACAGACTCATTTAGACACAGTAATTCACGACACAGAA 1020
Qy 1087 AAAGAAATCTAGTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAATCCATCA 1146
Db 1021 AAAGAAATCTAGTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAATCCATCA 1080
```

```
Qy 1147 ATGAGAAGACTTTTTCAGCTACTATCTCATTTGAAACTGGCCCTTTAAAAAGGACGACAG 1206
Db 1081 ATGAGAAGACTTTTTCAGCTACTATCTCATTTGAAACTGGCCCTTTAAAAAGGACGACAG 1140
Qy 1207 TTGAAGGAATTCATATGAAGACATTTCTAAACTTGCTCAGTTTGAAAAATTTGCTTAACA 1266
Db 1141 TTGAAGGAATTCATATGAAGACATTTCTAAACTTGCTCAGTTTGAAAAATTTGCTTAACA 1200
Qy 1267 ATACATTAAAACTTAAAGCATCAAAATTTGTTGTCGCAAGGACATTATGAGACTTACTG 1326
Db 1201 ATACATTAAAACTTAAAGCATCAAAATTTGTTGTCGCAAGGACATTATGAGACTTACTG 1260
Qy 1327 TGTAGGGTATATCTTTTGTATAAAAACAACAGGTTTTTGAAAAATTTACTGTATAGTT 1386
Db 1261 TGTAGGGTATATCTTTTGTATAAAAACAACAGGTTTTTGAAAAATTTACTGTATAGTT 1320
Qy 1387 GTTCAGCTAAACTTTTGAGAAGAAATTTAATTTATGTCATCAGGATCAAACTATGTAATT 1446
Db 1321 GTTCAGCTAAACTTTTGAGAAGAAATTTAATTTATGTCATCAGGATCAAACTATGTAATT 1380
Qy 1447 TTGTCCTTGTATTTT 1462
Db 1381 TTGTCCTTGTATTTT 1396
RESULT 2
CR618602 1339 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0DB008YG18 of Neuroblastoma Cot
DEFINITION 10-normalized of Homo sapiens (human).
ACCESSION CR618602 GI:50499409
VERSION CR618602.1 GI:50499409
KEYWORDS HTC; CNSLT_CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1339)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Faraday Avenue
Genoscope.
REFERENCE 2 (bases 1 to 1339)
AUTHORS Direct Submission
TITLE Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
1. .1339
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon.9606"
/clone="CS0DB008YG18"
/tissue_type="Neuroblastoma Cot 10-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 78.5%; Score 1199.4; DB 4; Length 1339;
Best Local Similarity 92.8%; Pred. No. 7.1e-264;
Matches 1311; Conservative 0; Mismatches 1; Indels 101; Gaps 1;
Qy 27 AGAAGTGGGTCCAGAAAGTGCCTGCGCTGCGTCCCATGGGAAAGTCGGATTTCTT 86
Db 28 AGTACTGGGTCCAGAAAGTGCCTGCGCTGCGTCCCATGGGAAAGTCGGATTTCTT 87
```

QY	87	ACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAGGGGCTGCCAAGACTACGCTGGTAT	146
Db	88	ACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAGGGGCTGCCAAGACTACGCTGGTAT	147
QY	147	TGCCAGATGTGCGCAGAAGCAGTCCCGGACGAGAAATGGCTTTAAAGTGTCAATTTGATGTCC	206
Db	148	TGCCAGATGTGCGCAGAAGCAGTCCCGGACGAGAAATGGCTTTAAAGTGTCAATTTGATGTCC	207
QY	207	GAATCTCATCAGAGACAATATTTGCTGGCTTCAGAAAATCTCAGCAGTCTTATGATGATAT	266
Db	208	GAATCTCATCAGAGACAATATTTGCTGGCTTCAGAAAATCTCAGCAGTCTTATGATGATAT	267
QY	267	TTTTTCAGAGGAAATCCGAAATGACATTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAA	326
Db	268	TTTTTCAGAGGAAATCCGAAATGACATTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAA	327
QY	327	AGGGTCCCAACACACATTTGCTTACAAACGAATACATCAGCCACCGAGAGACATCCACATG	386
Db	328	AGGGTCCCAACACACATTTGCTTACAAACGAATACATCAGCCACCGAGAGACATCCACATG	387
QY	387	AATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGTGTGCTGGCAGAGAGCGCTTG	446
Db	388	AATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGTGTGCTGGCAGAGAGCGCTTG	447
QY	447	TGCAAGTGGCAGACACCAAAAGCGTGGTATATTTTCACTACATAGCAGGACCCAGAA	506
Db	448	TGCAAGTGGCAGACACCAAAAGCGTGGTATATTTTCACTACATAGCAGGACCCAGAA	507
QY	507	ACTATCCGCGGCACTGGAACTGGAGAAAAGAAAAGCAGGACCTTGATGATGAAGAA	566
Db	508	ACTATCCGCGGCACTGGAACTGGAGAAAAGAAAAGCAGGACCTTGATGATGAAGAA	567
QY	567	AAAACTGCCAAATTTATGAAGACGAAGTGAGAAGAGCGCTCGAAAGGGAAGAAACAGGAG	626
Db	568	AAAACTGCCAAATTTATGAAGACGAAGTGAGAAGAGCGCTCGAAAGGGAAGAAACAGGAG	627
QY	627	GTCCCTACTTTTACGGAATTAAGCAGAGAAAATGATGAAGAGAAAGTCACTTTAATTTG	686
Db	628	GTCCCTACTTTTACGGAATTAAGCAGAGAAAATGATGAAGAGAAAGTCACTTTAATTTG	687
QY	687	AGTAAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCAACTCAAGTACTCTGGGACCG	746
Db	688	AGTAAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCAACTCAAGTACTCTGGGACCG	747
QY	747	AGTGCACTGAAGACGATAGGAAGTTTCAGCATCAGTGAACAGAAAAGAAATCTTCCAGAGC	806
Db	748	AGTGCACTGAAGACGATAGGAAGTTTCAGCATCAGTGAACAGAAAAGAAATCTTCCAGAGC	807
QY	807	TCAACTCAGTCTTAAGAAAAGAAAAGAAAATCTGCACTGATGAATCATGGAGATT	866
Db	808	TCAACTCAGTCTTAAGAAAAGAAAAGAAAATCTGCACTGATGAATCATGGAGATT	867
QY	867	GAAGAGAAAAGAAAAGAACTGCCGAAACAGACTACTGCTCAGCCTCAAAATTTATTTG	926
Db	868	GAAGAGAAAAGAAAAGAACTGCCGAAACAGACTACTGCTCAGCCTCAAAATTTATTTG	927
QY	927	AAAAATTATTAACCAAGAACTGGGAGAGAAATATCATTAAGAAAAGGCTATTGTTAAGGAA	986
Db	928	AAAAATTATTAACCAAGAACTGGGAGAGAAATATCATTAAGAAAAGGCTATTGTTAAGGAA	987
QY	987	GTAAATTGACAAATATACAGCTGTTGTGAAGATGATTTCTGGAGACAAGCTGAACATT	1046
Db	988	GTAAATTGACAAATATACAGCTGTTGTGAAGATGATTTCTGGAGACAAGCTGAACATT	1047
QY	1047	GACCAGACTATTTAGAGACAGTAATTTCCAGACACCAAGAAAAGAAATTTCTAGTTTTAAAT	1106
Db	1048	GACCAGACTATTTAGAGACAGTAATTTCCAGACACCAAGAAAAGAAATTTCTAGTTTTAAAT	1084
QY	1107	GGAGGCTACAGAGAAATGAAGGTACCCCTAGAAATCCATCAATGAGAAGACTTTTTCAGCT	1166
Db	1085	-----	1084

QY	1167	ACTATCGTCATTTGAAACTGGCCCTTTTAAAGGACGACAGAGTTGAAGAAATTCATATGAA	1226
Db	1085	-----GGCCCTTTTAAAGGACGACAGAGTTGAAGAAATTCATATGAA	1126
QY	1227	GACATTTCTAAACTTGGCTGAGTTTGAATAATTTGTTAACATACATTAATAATCTTAAAGC	1286
Db	1127	GACATTTCTAAACTTGGCTGAGTTTGAATAATTTGTTAACATACATTAATAATCTTAAAGC	1186
QY	1287	ATCAAAATGGTGTTCGCAAGGCATTATGAGACTCTACTGTGTAGGTATATTTCTTTT	1346
Db	1187	ATCAAAATGGTGTTCGCAAGGCATTATGAGACTCTACTGTGTAGGTATATTTCTTTT	1246
QY	1347	TATAAACAAACACAGTTTTTTGAAAAATTTACTGTATAGTTTTCAGCTAAACTTTTGAGAA	1406
Db	1247	TATAAACAAACACAGTTTTTTGAAAAATTTACTGTATAGTTTTCAGCTAAACTTTTGAGAA	1306
QY	1407	GAATTTAATATGCTCTCATGAGGTATCAAACTA	1439
Db	1307	GAATTTAATATGCTCTCATGAGGTATCAAACTA	1339

RESULT 3			
AY609991	1543 bp	mRNA	linear
LOCUS	Sus scrofa clone Clu_5302.scr.msk.pl.Contig4	HTC	31-JAN-2005
DEFINITION	Sus scrofa clone Clu_5302.scr.msk.pl.Contig4	HTC	31-JAN-2005
ACCESSION	AY609991		
VERSION	AY609991.1	GI:52351561	
KEYWORDS	HTC.		
SOURCE	Sus scrofa (pig)		
ORGANISM	Sus scrofa		
REFERENCE	1 (bases 1 to 1543)		
AUTHORS	Jorgensen, F.G., Hobolth, A., Hornshoj, H., Bendixen, C., Fredholm, M. and Schierup, M.H.		
TITLE	Comparative analysis of protein coding sequences from human, mouse and the domesticated pig		
JOURNAL	(er) BMC Biol. 3 (1), 2 (2005)		
PUBMED	15679890		
REFERENCE	2 (bases 1 to 1543)		
AUTHORS	Hornshoj, H., Bendixen, C. and Panitz, P.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-Apr-2004) Animal Breeding and Genetics, Danish Institute of Agricultural Sciences, Research Centre Foulum, Postbox 50, Tjele DK-8830, Denmark		
FEATURES	Location/Qualifiers		
source	1..1543		
	/organism="Sus scrofa"		
	/mol_type="mRNA"		
	/db_xref="taxon:9823"		
	/clone="Clu_5302.scr.msk.pl.Contig4"		
misc_feature	1..1543		
	/note="similar to NM_012311.2 Homo sapiens KIN, antigenic determinant of recA protein homolog (mouse) (KIN)"		
ORIGIN			
Query Match	77.8%	Score 1189.2;	DB 4; Length 1543;
Best Local Similarity	89.3%	Pred. No. 1.6e-261;	
Matches 1314;	Conservative 0;	Mismatches 152;	Indels 5; Gaps 3;
QY	55	CCGTGCTGCCATGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCA	114
Db	25	CTGTGCTCCGATGGGAAGTCGGATTTTCTTACCCGNAAGCCATCGCCACAGGATCA	84
QY	115	AGTCCAAGGGCTCGAGAAGCTACCTGTATTTGCCAGATGTGCCAGAAGCAGTGCCTGG	174
Db	85	AATCCAAGGGCTTCAGAAGCTCGCTGCTGTAATTCAGATGTGCCAGAAGCAGTGCCTGG	144
QY	175	ACGAGAAATGGCTTTAAGTGTCTATTTGTCGGAATCTCATCAGAGACAATCTTTCCTGG	234
Db	145	ACGAGAAATGGCTTTAAGTGTCTATTTGTCGGAATTCGCAATTCGATCAGAGACAATCTTTCCTGG	204

Df	421	TACATAGACAGGACCCAGAAACTATCCCGCGCACTGGAACTGGAGAGAAAGAAAG	480
Qy	546	CAGGACCTTGATGATGAGAAAAAATCTGCCAAATTTATTGAAGACAAAGTGAGAGAGGC	605
Df	481	CAGGACCTTGATGATGAGAAAAAATCTGCCAAATTTATTGAAGACAAAGTGAGAGAGGC	540
Qy	606	CTGGAAGGAGGACAGGAGTCCCTACTTTTTCGGAATTAAGCAGAGAAATGATGAA	665
Df	541	CTGGAAGGAGGAGGAGGAGTCCCTACTTTTTCGGAATTAAGCAGAGAAATGATGAA	600
Qy	666	GAGAAAGTCAAGTTTAAATTTAGTAAGAGGAGCATGTAGCTCATCCGGAGCAACATCTTCC	725
Df	601	GAGAAAGTCAAGTTTAAATTTAGTAAGAGGAGCATGTAGCTCATCCGGAGCAACATCTTCC	660
Qy	726	AAGTCAAGTACTCTCGGACCGAGTGCATGGAAGACGATAGGAAGTTCAGCATCAGTGAAA	785
Df	661	AAGTCAAGTACTCTCGGACCGAGTGCATGGAAGACGATAGGAAGTTCAGCATCAGTGAAA	720
Qy	786	CGAAAGAAATCTTCCAGAGCTCACTCAGTCTAAGAAAGAGAGAAAGAAATCTGCA	845
Df	721	CGAAAGAAATCTTCCAGAGCTCACTCAGTCTAAGAAAGAGAGAAAGAAATCTGCA	780
Qy	846	CTGGATGAAATCATGGAGATTGAAGAGGAGAAAGAAAGAACTGCCCGAACAGACTACTGG	905
Df	781	CTGGATGAAATCATGGAGATTGAAGAGGAGAAAGAAAGAACTGCCCGAACAGACTACTGG	840
Qy	906	CTACAGCTGAAATTTATTTGTGAAAATTAACCAAGAACTGGGAGAGAAATATCATAG	965
Df	841	CTACAGCTGAAATTTATTTGTGAAAATTAACCAAGAACTGGGAGAGAAATATCATAG	900
Qy	966	AAAAGGCTATTTTAAAGAGTAAATTCAGCAATATACAGCTGTTGTGAAGATGATGAT	1025
Df	901	AAAAGGCTATTTTAAAGAGTAAATTCAGCAATATACAGCTGTTGTGAAGATGATGAT	960
Qy	1026	TCTGAGAGCAAGCTGAAACTTGCACAGACTCATTTTAGAGACAGTAATTTCCAGCACGAGA	1085
Df	961	TCTGAGAGCAAGCTGAAACTTGCACAGACTCATTTTAGAGACAGTAATTTCCAGCACGAGA	1020
Qy	1086	AAAAGAAATCTAGTTTTTAAATGAGGCTACAGAGAAATGAAGTACCCTAGAAATCCATC	1145
Df	1021	AAAAGAAATCTAGTTTTTAAATGAGGCTACAGAGAAATGAAGTACCCTAGAAATCCATC	1080
Qy	1146	AATGAGAGACTTTTTCAGCTACTATCGTCAATGGAAGTGGCCCTTTAAAGAGCGAGA	1205
Df	1081	AATGAGAGACTTTTTCAGCTACTATCGTCAATGGAAGTGGCCCTTTAAAGAGCGAGA	1140
Qy	1206	GTTGAAGGAAATTCATATGAGACATTTCTAAACTTTGGCTGA	1247
Df	1141	GTTGAAGGAAATTCATATGAGACATTTCTAAACTTTGGCTGA	1182

RESULT 6
AL522388
LOCUS
DEFINITION
AL522388 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens
cDNA clone CS0DB008Y18 5-PRIME, mRNA sequence.
ACCESSION
AL522388
VERSION
AL522388.3 GI:45697719
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 965)
Li W.B., Gruber C., Jessee J. and Polayes D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 13, 2001 this sequence version replaced gi:31040660.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 5543.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna/?s=CS0DB008Y18&c=5543.r.

Location/Qualifiers
1. .965
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DB008Y18"
/tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match	58.9%;	Score 900;	DB 1;	Length 965;
Best Local Similarity	99.5%;	Pred. No. 2.9e-195;		
Matches	933;	Conservative	1;	Mismatches 3; Gaps 3;
Qy	28	GAACTGGGGTCCAGAAAGTGATGCTGCCGTGGTGGCAAGTTCGGATTTCTTA	87	
Df	29	GATCTGGGGTCCAGAAAGTGATGCTGCCGTGGTGGCAAGTTCGGATTTCTTA	88	
Qy	88	CTCCCAAGGCTATGCCCAACAGGATCAAGTCCCAAGGGCTGCAGAGCTAGCTGGTAT	147	
Df	89	CTCCCAAGGCTATGCCCAACAGGATCAAGTCCCAAGGGCTGCAGAGCTAGCTGGTAT	148	
Qy	148	GCCAGATGTGCCAGAGCAGTGGCGGACGAGAAATGCTTTAAGTGTCATTTATGTCGG	207	
Df	149	GCCAGATGTGCCAGAGCAGTGGCGGACGAGAAATGCTTTAAGTGTCATTTATGTCGG	208	
Qy	208	AATCTCATCAGAGACAACTATTTGCTGCTTCAGAAAATCTCTCAGAGTTATGATTA	267	
Df	209	AATCTCATCAGAGACAACTATTTGCTGCTTCAGAAAATCTCTCAGAGTTATGATTA	268	
Qy	268	TTTTCAGAGGAAATCCGAAATGACCTTTCTAGAACTTTCTCAGAGAGCGCTTTGGCACTAAA	327	
Df	269	TTTTCAGAGGAAATCCGAAATGACCTTTCTAGAACTTTCTCAGAGAGCGCTTTGGCACTAAA	328	
Qy	328	GGGTCCCAACAACTATTTGCTGCTTCAGAAAATCTCTCAGAGAGCGCTTTGGCACTAAA	387	
Df	329	GGGTCCCAACAACTATTTGCTGCTTCAGAAAATCTCTCAGAGAGCGCTTTGGCACTAAA	388	
Qy	388	ATGCCACTCAGTGGGAACTCTGACTGATTTTCTAAGTGGCTGGGAGAGAGCGCTTGT	447	
Df	389	ATGCCACTCAGTGGGAACTCTGACTGATTTTCTAAGTGGCTGGGAGAGAGCGCTTGT	448	
Qy	448	GCAAGTGGAGAGACACCAAGGCTGATATTTCTAGTACATAGACAGGAGCCAGGAA	507	
Df	449	GCAAGTGGAGAGACACCAAGGCTGATATTTCTAGTACATAGACAGGAGCCAGGAA	508	
Qy	508	CTATCCCGCGCAACTGGAACTGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG	567	
Df	509	CTATCCCGCGCAACTGGAACTGGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG	568	
Qy	568	AAACTGCCAAATTTATTGAAGAGCAAGTGAAGAGAGCGCTTGAAGAGAGAGAGAGAG	627	
Df	569	AAACTGCCAAATTTATTGAAGAGCAAGTGAAGAGAGCGCTTGAAGAGAGAGAGAGAG	628	
Qy	628	TCCCTACTTTTACGAAATTAAGCAGAGAAATGATGAAGAGAGAGAGAGAGAGAGAG	687	
Df	629	TCCCTACTTTTACGAAATTAAGCAGAGAAATGATGAAGAGAGAGAGAGAGAGAGAG	688	
Qy	688	GTAAGAGAGAGATGTAGCTCATCCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGA	747	
Df	689	GTAAGAGAGAGATGTAGCTCATCCGGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGA	748	

QY	748	GTGCACTGAAGAGGTAGGAGTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAGCT	807
Db	749	GTGCACTGAAGAGGTAGGAGTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAGCT	807
QY	808	CAACTCAGTCTTAAGAAAAAGAAAAAG- AAATCTGCACTGGATGAATCATGAGAGATT	866
Db	808	CAACTCAGTCTTAAGAAAAAGAAAAAGAAATCTGCACTGGATGAATCATGAGAGATT	867
QY	867	GAAGAGG-AAAAGAAAAGTCCCGAAGAGAGTCTGCTACAGCTGAATATTGTT	925
Db	868	GAAGAGGAAAAGAAAAGTCCCGAAGAGAGTCTGCTACAGCTGAATATTGTT	927
QY	926	GAATATTATTAACCAAGAACTGGAGAGAAATATCAT	963
Db	928	GAATATTATTAACCAAGAACTGGAGAGAAATATCAT	965
RESULT 7			
LOCUS			
DEFINITION			
ILLUMIGEN MCQ_40291 Katze_MMPB2 Macaca mulatta cDNA Clone			
IBIUM:27173 5' similar to Bases 5 to 950 highly similar to human			
KIN (Hs.123647), mRNA sequence.			
CO725903			
CO725903.1 GI:50707056			
EST.			
Macaca mulatta (rhesus monkey)			
Macaca mulatta			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
Cercopithecoidea; Cercopithecinae; Macaca.			
1 (bases 1 to 951)			
Magness,C.L., Fellin,P.C., Thomas,M.J., Korth,M.J., Agy,M.B.,			
Prolli,S.C., Fitzgibbon,M., Scherer,C.A., Miner,D.G., Katze,M.G. and			
Iadonato,S.P.			
Analysis of the Macaca mulatta transcriptome and the sequence			
divergence between Macaca and human			
Genome Biol. 6 (7), R60 (2005)			
15998449			
Contact: C. Magness			
Illumigen Biosciences Inc.			
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA			
Tel: 2063780400			
Fax: 2063780408			
Email: cmagness@illumigen.com			
Sequenced on 2004.05.28. 723 Q20 bases. Library Preparation: Prof.			
Michael Katze Lab at University of Washington DNA Sequencing:			
Illumigen Biosciences Inc. For further information, see			
http://www.macaque.org			
PCR Primers			
FORWARD: CCTCACTAAGGGAAACAAA			
BACKWARD: CACTATAGGCGAATTGGGTA			
Insert Length: 951 Std Error: 0.00			
Plate: CL080321 row: C column: 12			
Seq primer: CCTCACTAAGGGAAACAAA			
POLYA=No.			
Location/Qualifiers			
1. .951			
/organism="Macaca mulatta"			
/mol_type="mRNA"			
/strain="Indian"			
/db_xref="taxon:9544"			
/clone="IBIUM:27173"			
/sex="male"			
/tissue_type="blood"			
/cell_type="PBMC"			
/dev_stage="adult"			
/lab_host="Electromax DH10B"			
/clone_lib="Katze MMPB2"			
/note="Vector: pDONR 222; Site 1: BsrG I; Site 2: BsrG I;			
Created from CloneMiner cDNA Library Construction kit			
(catalog #18249-029)"			
FEATURES			
source			
Query Match			
Best Local Similarity 56.2%; Score 859.2; DB 7; Length 951;			
Matches 901; Conservative 96.3%; Pred. No. 6.4e-186;			
Mismatches 0; Indels 2; Gaps 2;			
QY	28	GAACTGGGGTCCAGAAAGTGATCGCTGCGTGGTCCCATGGGGAAGTCGGATTTCTTA	87
Db	15	GCACTAGGGTCCAGAAAGTGACCGCTGCGTGGTCCCATGGGGAAGTCGGATTTCTTA	74
QY	88	CTCCCAAGGCTATCGCCCAACAGGATCAAGTCCAAAGGGCTGCAGAGCTTACCTCGTATT	147
Db	75	CTCCCAAGGCTATCGCCCAACAGGATCAAGTCCAAAGGGCTGCAGAGCTTACCTCGTATT	134
QY	148	GCAGATGTGCCAGAAAGCAGTGCCTGGGACGAGATGGCTTTAAAGTGTCTATTGTCTCG	207
Db	135	GCAGATGTGCCAGAAAGCAGTGCCTGGGACGAGATGGCTTTAAAGTGTCTATTGTCTCG	194
QY	208	AATCTCATCAGAGACAACTATTGTGGCTTCAGAAAAATCTCTCAGCAGTTTATGGATTATT	267
Db	195	AATCTCATCAGAGACAACTATTGTGGCTTCAGAAAAATCTCTCAGCAGTTTATGGATTATT	254
QY	268	TTTCAGAGAAATTCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCAGCTAAAA	327
Db	255	TTTCAGAGAAATTCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCAGCTAAAA	314
QY	328	GGGTCCCAACAACTATTGTCTACAAAGATACATCAGCCAGAGACATCCACATCA	387
Db	315	GGGTCCCAACAACTATTGTCTACATAGATACATCAGCCAGAGACATCCACATCA	374
QY	388	ATGCCACTCAGTGGGAACTCTGACTGATTTTAAAGTGGCTGGGACAGAGAGGTTGT	447
Db	375	ATGCCACTCAGTGGGAACTCTGACTGATTTTAAAGTGGCTGGGACAGAGAGGTTGT	434
QY	448	GCAAGTGACAGACACCAAAAGGCTGTATTTACAGTACATAGACAGGACCCAGAAA	507
Db	435	GCAAGTGATGAGACACCAAAAGGCTGTATTTACAGTACATAGACAGGACCCAGAAA	494
QY	508	CTATCCGCGGCAACTGGAACTGGAGAAAGAAAGAAAGCAGGACCTTTGATGAGAGAA	567
Db	495	CTATCCGCGGCAACTGGAACTGGAGAAAGAAAGAAAGCAGGACCTTTGATGAGAGAA	554
QY	568	AAACTGCCAAATTTATTGAAGACAGTGAAGAGGCTCGAAAGGAAAGGAAACAGGAGG	627
Db	555	AAACTGCCAAATTTATTGAAGAAACAGTGAAGAGGCTCGAAAGGAAAGGAAACAGGAGG	614
QY	628	TCCTACTTTTACGGAATTAAGCAGAGAAATGATGAAGAGAAAGTCAGCTTTAATTGA	687
Db	615	CCCTACTTTTACGGAATTTAAGCAGAGAAATGATGAAGAGAAAGTCAGCTTTAATTGA	674
QY	688	GTAAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGA	747
Db	675	GTAAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCAAAGTCAAGTACTCTGGGACCGA	734
QY	748	GTGCACTGAAGACGATAGGAAGTTACAGCATCAGTGAAGAAAGAAAGAAATTTTCCAGAGCT	807
Db	735	GTGCACTGAAGACGATAGGAAGTTACAGCATCAGTGAAGAAAGAAAGAAATTTTCCAGAGCT	794
QY	808	CAACTCAGTCTAAAGAAAAGAAAGAAATCTGCATCTGGATGAATCATCGAGAA-TT	866
Db	795	CACCTCAGTCTCAAAAAGAAAGAAAGAAATCTGCATCTGGATGAATCATCGAGAAATTT	854
QY	867	GAAGAGGAAAGAAAGAAAGTCCCGAAGACAGACTACTGGCTCAGAGCTGAAATTT-ATTGT	925
Db	855	GAAGAGGAAAGAAAGAAAGTCCCGAAGACAGACTACTGGCTCAGAGCTGAAATTTATTGT	914
QY	926	GAATATTATTAACCAAGAACTGGGAGAGAAATATCA	961
Db	915	GAATATTATTAACCAAGAACTGGGAGAGAAATATCA	950
RESULT 8			
C0726806			

LOCUS C0726806 988 bp mRNA linear EST 27-JUL-2004
DEFINITION ILLUMIGEN_MCO_48930 Katze_MMD Macaca mulatta cDNA clone
IBUW:27517 5'-similar to Bases 5 to 967 highly similar to human
KIN (Hs.123647), mRNA sequence.
ACCESSION C0726806
VERSION C0726806.1 GI:50708873
KEYWORDS EST.
SOURCE Macaca mulatta (rhesus monkey)
ORGANISM Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Cercopithecoidea; Cercopithecinae; Macaca.
REFERENCE 1 (bases 1 to 988)
AUTHORS Magness,C.L., Fellin,P.C., Thomas,M.J., Korth,M.J., Agy,M.B.,
Prohl,S.C., Fitzgibbon,M., Scherer,C.A., Miner,D.G., Katze,M.G. and
Iadonato,S.P.
TITLE Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human
JOURNAL Genome Biol. 6 (7), R60 (2005)
PUBMED 15998449
COMMENT Contact: C. Magness
Illumigen Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagness@illumigen.com
Sequenced on 2004.07.02. 706 Q20 bases. Library Preparation: Prof.
Michael Katze Lab at University of Washington DNA Sequencing:
Illumigen Biosciences Inc. For further information, see
http://www.macaque.org
PCR Primers
FORWARD: CCCTCACTAAGGGACAAAA
BACKWARD: CACTATAGGCGAATGGTA
Insert Length: 988 Std Error: 0.00
Plate: CL000405 row: C column: 07
Seq primer: CCCTCACTAAGGGACAAAA
POLYA=Yes location/Qualifiers
i..988
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="IBUW:27517"
/sex="male"
/dev_stage="adult"
/lab_host="Electronmax DH10B"
/clone_lib="Katze_MMD"
/note="Organ: duodenum; Vector: pDONR 222; Site 1: BsrG I;
Site 2: BsrG I; Created from Clonewiner cDNA Library
Construction kit (catalog #18249-029)"

ORIGIN

Query Match 55.6%; Score 849; DB 7; Length 988;
Best Local Similarity 94.0%; Pred. No. 1.4e-183;
Matches 915; Conservative 0; Mismatches 55; Indels 3; Gaps 3;
QY 39 CAGAAAGTGATCGCTGCGGTGTCGCATGGGGAAGTCGGATTTCTTACTCCCAAGGCT 98
DB 5 CAGAAAGTGACCGCTGCGGTGTCGCATGGGGAAGTCGGATTTCTTACTCCCAAGGCT 64
QY 99 ATCGGCAACAGAGATCAAGTCCAAAGGGCTGCGAGAGCTACGCTGGTATGCCAGATGTGC 158
DB 65 ATCGGCAACAGAGATCAAGTCCAAAGGGCTGCGAGAGCTACGCTGGTATGCCAGATGTGC 124
QY 159 CAGAAAGTGTGCGGACGAGAAATGGCTTTAAGTGTCAATTGATGTGCGAAATCTCATCAG 218
DB 125 CAGAAAGTGTGCGGACGAGAAATGGCTTTAAGTGTCAATTGATGTGCGAAATCTCATCAG 184
QY 219 AGACAACATATTGCTGGCTTCAGAAAAATCCTCAGCAGTTTATGAGTATTTTTCAGAGGAA 278
DB 185 AGACAACATATTGCTGGCTTCAGAAAAATCCTCAGCAGTTTATGAGTATTTTTCAGAGGAA 244

QY 279 TTCGAAATGACTTTCTAGAACTTCTCAGGAGAGCGTTTGGCACTAAAGGGTCCACAC 338
DB 245 TTCGAAATGACTTTCTAGAACTTCTCAGGAGAGCGTTTGGCACTAAAGGGTCCACAC 304
QY 339 AACATTGTCTACAAAGAAATACATCAGCCACCGAGAGCACATCCACATGAATGCCACTCAG 398
DB 305 AACATTGTCTACATAGTACATCAGCCACCGAGAGCACATCCATATGAATGCAACTCAG 364
QY 399 TGGAAACTCTGACTGATTTTAAAGTGGCTGGCGAGAGAGGCTTGTGCAAAAGTGAC 458
DB 365 TGGAAACTCTGACCGATTTTAAAGTGGCTGGCGAGAGAGGCTTGTGCAAAAGTGAT 424
QY 459 GAGACACCAAAAGGCTGTATATTACGTACATACAGAGGACCCAGAAACTATCCGCGG 518
DB 425 GAGACACCAAAAGGCTGTATATTACGTACATACAGAGGACCCAGAAACTATTCGCGG 484
QY 519 CAACCTGGAACCTGGAGAAAAAGAAAAAGCAGGACCTTGATGATGAAGAAAAAACTGCCAAA 578
DB 485 CAACCTGGAACCTGGAGAAAAAGAAAAAGCAGGACCTTGATGATGAAGAAAAAACTGCCAAA 544
QY 579 TTTATTGAAGCAAGTGAAGAGGCTTGAAGGAAAGGAAACAGGAGGTCCCTACTTTT 638
DB 545 TTTATTGAAGCAACAGTGAGAAGGCTTGAAGGAAAGGAAACAGGAGGCTCCCTACTTTT 604
QY 639 ACGAATTTAAGCAGAGAAATGATGAAGAGAACTCAGTTTAATTTGAGTAAAGGAGCA 698
DB 605 ACGAATTTAAGCAGAGAAATGATGAAGAGAACTCAGTTTAATTTGAGTAAAGGAGCA 664
QY 699 TGTAAGTCTATCCGAGCAACATCTTCAAGTCAAGTACTCTGGGACCGAGTGCATGAAG 758
DB 665 TGCAGTTTATCCGAGCAACGCTCTCCAGTCAAGTACTCTGGGACCGAGTGCATGAAG 724
QY 759 ACGATAGAGAGTTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 818
DB 725 ACAATAAGAGTTTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 784
QY 819 AAGAAAGAGAGAAAGAAATCTGCACTGCAATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 878
DB 785 AAGAAAGAGAGAAAGAAATCTGCACTGCAATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 843
QY 879 AAAAGAACTGCCGGAACAGACTACTGGCTACAGCTGCAAGTCAAGTCAAGTCAAGTCAAGT 938
DB 844 AAAAGAACTGCCGGAACAGACTACTGGCTACAGCTGCAAGTCAAGTCAAGTCAAGTCAAGT 902
QY 939 AAGAACTGGGAGAGAAATATCATAGAAAGGCTATTTGTTAAGGAGTAAATTTGACAAA 998
DB 903 AA-AAACTGGGAGAGAAATATCCAAAAAGGGAACCGTAAAGGAGATATTTTGGAAA 961
QY 999 TATACAGCTGTTG 1011
DB 962 AATATAAGCGCTG 974

RESULT 9
LOCUS AL558811
DEFINITION AL558811 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
Homo sapiens cDNA clone CS0DJ015Yil9 5-PRIME, mRNA sequence.
ACCESSION AL558811
VERSION AL558811.3 GI:46184200
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 846)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 15, 2001 this sequence version replaced gi:31280609.
Contact: Genoscope
Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 5543.r

For more information about this cluster, see
 http://www.genoscope.cns.fr/cdna?b=CS0DU015AE100P1&c=5543.r.

FEATURES

Location/Qualifiers
 1. .846
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DU015Y119"
 /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
 /cell_line="JURKAT"
 /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
 10-NORMALIZED"
 /notes="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 54.4%; Score 831; DB 1; Length 846;
 Best Local Similarity 99.6%; Pred. No. 1.9e-179;
 Matches 842; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 67 TGGGAGTCCGATTTCTTACTCCCAAGGCTATCGCCACAGATCAAGTCCAAAGGGC 126
 DB 1 TGGGAGTCCGATTTCTTACTCCCAAGGCTATCGCCACAGATCAAGTCCAAAGGGC 60

QY 127 TGCAGAGCTACGCTGGTATTGTCAGATGTGCCAGAGCAGTCCCGGACGAGAATGGCT 186
 DB 61 TGCAGAGCTACGCTGGTATTGTCAGATGTGCCAGAGCAGTCCCGGACGAGAATGGCT 120

QY 187 TTAAGTGTCAATGTATGTCGAAATCTCATCAGAGCAACTATTGCTGGCTTCAGAAATC 246
 DB 121 TTAAGTGTCAATGTATGTCGAAATCTCATCAGAGCAACTATTGCTGGCTTCAGAAATC 180

QY 247 CTGACAGTGTATGGATTTATTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTCA 306
 DB 181 CTCAGCAGTTATGGATTTATTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTCA 240

QY 307 GGAGACGCTTTGGCACTAAAGGGTCCACAAACATTTGCTACAAAGATATACATCAGCC 366
 DB 241 GGAGACGCTTTGGCACTAAAGGGTCCACAAACATTTGCTACAAAGATATACATCAGCC 300

QY 367 ACCGAGAGCATCCCATGATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGT 426
 DB 301 ACCGAGAGCATCCCATGATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGT 360

QY 427 GGCTGGCAGAGAGGCTTGGCAAGTGGACGAGACCAAAAGGCTGGTATATTTCAGT 486
 DB 361 GGCTGGCAGAGAGGCTTGGCAAGTGGACGAGACCAAAAGGCTGGTATATTTCAGT 420

QY 487 ACATAGACAGGACCCAGAAATCTATCCCGCGCACTGGAACTGGAGAGAGAGAGAGC 546
 DB 421 ACATAGACAGGACCCAGAAATCTATCCCGCGCACTGGAACTGGAGAGAGAGAGAGC 480

QY 547 AGGACCTTGATGAGC 606
 DB 481 AGGACCTTGATGAGC 540

QY 607 TGGAGGAG 666
 DB 541 TGGAGGAG 600

QY 667 AGAAGTCACTTTTAAATTTAGTAAAGAGAGATGTAGCTCATCCGGAGCAACATCTTCCA 726
 DB 601 AGAAGTCACTTTTAAATTTAGTAAAGAGAGATGTAGCTCATCCGGAGCAACATCTTCCA 660

QY 727 AGTCAAGTACTCTGGACCGAGTGCACCTGAAGACGATAGGAAGTTTCAGCATCAGTGAAC 786
 DB 661 AGTCAAGTACTCTGGACCGAGTGCACCTGAAGACGATAGGAAGTTTCAGCATCAGTGAAC 719

QY 787 GAAAGAAATCTTCCAGAGCTCACTAGTCTTAAAGAAAAGAAAAGAAAAGAAAATCTGCAC 846
 DB 720 GAAAGAAATCTTCCAGAGCTCACTAGTCTTAAAGAAAAGAAAAGAAAAGAAAATCTGCAC 779

QY 847 TCGATGAATCATGAGATTGAAGAGGAGAAAAGAAAAGAACTCCCGAAGCAGACTACTGGC 906
 DB 780 TCGATGAATCATGAGATTGAAGAGGAGAAAAGAAAAGAACTCCCGAAGCAGACTACTGGC 839

QY 907 TACAGCC 913
 DB 840 TACAGCC 846

RESULT 10
 BQ221694
 LOCUS
 DEFINITION
 5', mRNA sequence.
 ACCESSION
 BQ221694
 VERSION
 BQ221694.1 GI:20403094
 EST.
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 REFERENCE
 1. (bases 1 to 882)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: DCTD/DRP/Gazdar
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.G.E. Consortium (LLNL)
 DNA sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM13321 row: 1 column: 07
 High quality sequence stop: 581.
 Location/Qualifiers
 1. .882
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6057750"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 68"
 /notes="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.8 kb. Library constructed by Life
 Technologies."

ORIGIN

Query Match 52.4%; Score 800.4; DB 3; Length 882;
 Best Local Similarity 99.6%; Pred. No. 1.9e-172;
 Matches 812; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 36 GTCCAGAAAGTGTATCGCTGCCGTGGTCCCATCGGGAAGTCGGAATTTTCTTACTCCCAAG 95
 DB 1 GTCCAGAAAGTGTATCGCTGCCGTGGTCCCATCGGGAAGTCGGAATTTTCTTACTCCCAAG 60

QY 96 GCTATGCCCAACAGGATCAAGTCCAGGGGCTCCAGAGCTACGCTGGTATTGCCAGATG 155
 DB 61 GCTATGCCCAACAGGATCAAGTCCAGGGGCTCCAGAGCTACGCTGGTATTGCCAGATG 120

DEFINITION	603391378F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5406499 5', mRNA sequence.
ACCESSION	BI862461
VERSION	BI862461.1 GI:16003208
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 842)
TITLE	NIH-MGC http://mgs.nci.nih.gov/
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.
	Email: cgaabs-r@mail.nih.gov
	Tissue Procurement: DCTP/DTP
	cDNA Library Preparation: Life Technologies, Inc.
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
	DNA sequencing by: Incyte Genomics, Inc.
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:
	http://image.lnl.gov
	Plate: LINL2036 Row: 1 Column: 20
	High quality sequence stop: 834.
FEATURES	Location/Qualifiers
source	1..842
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="IMAGE:5406499"
	/tissue_type="mammary adenocarcinoma, cell line"
	/lab_host="DH10B (phage-resistant)"
	/clone_lib="NIH MGC 87"
	/notes="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT-primed. Average insert size 1.383 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."
ORIGIN	
Query Match	48.7%; Score 744.6; DB 3; Length 842;
Best Local Similarity	98.8%; Pred. No. 1.2e-159;
Matches 813; Conservative	0; Mismatches 4; Indels 6; Gaps 6;
QY	27 AGAATCGGGTCCAGAAAGTGATCGTGGCC-GTGGTCCCATGGGAAGTCGGATTTCCT 85
DB	10 AGTACTGGGTCCAGAAAGTGATCGTCCGGTGGTCCCATGGGAAGTCGGATTTCCT 69
QY	86 TACTCCCAAGGCTATGCCAACAGGATCAAGTCCAGGGGCTGCAGAGCTACGCTGGTA 145
DB	70 TACTCCCAAGGCTATGCCAACAGGATCAAGTCCAGGGGCTGCAGAGCTACGCTGGTA 129
QY	146 TTCCCGAGTGTCCAGAGCAGTGC CGGGACGAGATGGCTTTAAGTGTCATTGTATGTC 205
DB	130 TTCCCGAGTGTCCAGAGCAGTGC CGGGACGAGATGGCTTTAAGTGTCATTGTATGTC 189
QY	206 CGAATCTCATCAGAGACAACATTATGTGGCTTCAGAAAATCCTCAGCAGTTATGGATTA 265
DB	190 CGAATCTCATCAGAGACAACATTATGTGGCTTCAGAAAATCCTCAGCAGTTATGGATTA 249
QY	266 TTTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTCAGAGAGCGCTTTGGCACTAA 325
DB	250 TTTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTCAGAGAGCGC-TTGGCACTAA 308
QY	326 AAGGGTCCACAACAATTCTCTACACGAATACATCAGCCACCGAGAGCACATCCACAT 385
DB	309 AAGGGTCCACAACAATTCTCTACACGAATACATCAGCCACCGAGAGCACATCCACAT 368
QY	386 GAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAGTGGCTGGGCAGAGAAGGCTT 445
DB	369 GAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAGTGGCTGGGCAGAGAAGGCTT 428
QY	446 GTGCAAGTGGGACGAGACACCAAAAGGCTGGTATATTCAGTACATAGACAGGACCACGA 505
DB	429 GTGCAAGTGGGACGAGACACCAAAAGGCTGGTATATTCAGTACATAGACAGGACCACGA 488
QY	506 AACTATCCCGCGCAACTGGAACTGGAGAAAAAGAAAAAGCAGGACCTTTGATGATGAAGA 565
DB	489 AACTATCCCGCGCAACTGGAACTGGAGAAAAAGAAAAAGCAGGACCTTTGATGATGAAGA 548
QY	566 AAAAACTGCCAAATTTATTTGAAGAGCAAGTGAAGAGGCGCTTGGAGGAAGAAACAGGA 625
DB	549 AAAAACTGCCAAATTTATTTGAAGAGCAAGTGAAGAGGCGCTTGGAGGAAGAAACAGGA 608
QY	626 GTGCCCTACTTTTACGGAATTAAGCAGAGAA-AAATGATCAAGAGAAAGTCAGTTTAATT 684
DB	609 GTGCCCTACTTTTACGGAATTAAGCAGAGAA-AAATGATCAAGAGAAAGTCAGTTTAATT 668
QY	685 TGAGTAAAGGAGCATGTAGTCTCATCGGAGCAACATCTTCCAAAGTC-AAATGATCTCTGGGA 743
DB	669 TGAGTAAAGGAGCATGTAGTCTCATCGGAGCAACATCTTCCAAAGTC-AAATGATCTCTGGGA 728
QY	744 CCGAGTGCCT- GAGACCATAGGAAGTTTCAGATCAGTGAACGAAAAAGAAATCTTCCCA 802
DB	729 CCGAGTGCCTGGAAGACCATAGGAAGTTTCAGATCAGTGAACGAAAAAGAAATCTTCCCA 788
QY	803 GAGCTCAACT-CAGTCTAAAGAAAAAGAAAAAGAAATCTGC 844
DB	789 GAGCTCAACTCCAGTCTTAAAGAAAAAGAAAAAGAAATCTGC 831
RESULT 14	
LOCUS	CO725073 936 bp mRNA linear EST 27-JUL-2004
DEFINITION	ILLUMIGEN_MQ_26923 Katze_MMBR Macaca mulatta cDNA clone
	IBIWI:26022 57, similar to Bases 105 to 936 highly similar to human
	KIN (Hs.123647), mRNA sequence.
ACCESSION	CO725073
VERSION	CO725073.1 GI:50705382
KEYWORDS	EST.
SOURCE	Macaca mulatta (rhesus monkey)
ORGANISM	Macaca mulatta
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Cercopitheidae; Cercopitheciniae; Macaca.
REFERENCE	1 (bases 1 to 936)
AUTHORS	Magness,C.L., Fellin,P.C., Thomas,M.J., Korth,M.J., Agy,M.B., Proll,S.C., Fitzgibbon,M., Scherer,C.A., Miner,D.G., Katze,M.G. and Iadonato,S.P.
TITLE	Analysis of the Macaca mulatta transcriptome and the sequence divergence between Macaca and human
JOURNAL	Genome Biol. 6 (7), R60 (2005)
PUBMED	15998449
COMMENT	Contact: C. Magness Illumigen Biosciences Inc. 2203 Airport Way S, Suite 450, Seattle, WA 98134, USA Tel: 2063780400 Fax: 2063780408 Email: cmagness@illumigen.com Sequenced on 2004.03.11. 732 Q20 bases. Library Preparation: Prof. Michael Katze Lab at University of Washington DNA Sequencing: Illumigen Biosciences Inc. For further information, see http://www.macaque.org
PCR Primers	
FORWARD:	CCCTCACTAAAGGAACAAAA
BACKWARD:	CACATATAGGCGCAATTTGGTA
Insert Length:	936 Std Error: 0.00
Plate:	CL000184 row: B column: 01
Seq primer:	CCCTCACTAAAGGAACAAAA
POLYA=yes.	Location/Qualifiers
FEATURES	1..936
source	/organism="Macaca mulatta"
	/mol_type="mRNA"
	/strain="Indian"


```

/db xref="taxon:9544"
/clone="IBIUM:26022"
/sex="female"
/dev stage="adult"
/clone lib="Katze NMBR"
/notes="Organ: brain; Vector: Uni-ZAP XR; Site 1: EcoR I;
Site 2: Xho I; Created from Stratagene ZAP-CDNA Synthesis
kit (catalog #200400) and ZAP-CDNA Gigapack III Gold
Cloning Kit (Catalog #200450)"

ORIGIN
Query Match      48.7%; Score 744.2; DB 7; Length 936;
Best Local Similarity 95.2%; Pred. No. 1.5e-159;
Matches 789; Conservative 0; Mismatches 38; Indels 2; Gaps 2;

Qy 275 GGAATTCGGAATGACATTTCTAGAACTTCTCAGGAGCGCTTTGGCACTAAAGGGTCCA 334
Db 105 GGAATTCGGAATGACATTTCTAGAACTTCTCAGGAGCGCTTTGGCACTAAAGGGTCCA 164

Qy 335 CAACAACATTTCTACAACGATACATCAGCCCGAGAGCACATCCACATGATGCCAC 394
Db 165 CAACAACATTTCTACAATGATGATCAGCCCGAGAGCACATCCATATGAATGCAAC 224

Qy 395 TCAGTGGGAACTCTGACTGATTTTACTTAAGTGTCTGGCAGAGAGGCTTTGTCAAAGT 454
Db 225 TCAGTGGGAACTCTGACCGATTTTACTTAAGTGTCTGGCAGAGAGGCTTTGTCAAAGT 284

Qy 455 GGACGAGACACCAAAAGGCTGTATATTCAGTACATAGACAGGACCCAGAAACTATCCG 514
Db 285 GGATGAGACACCAAAAGGCTGTATATTCAGTACATAGACAGGACCCAGAAACTATTCG 344

Qy 515 CCGGCAACTGGAACCTGGAGAAAAGAAAAGAACAGCAGGACCTTGATGAGAAAACCTGC 574
Db 345 CCGGCAACTGGAACCTGGAGAAAAGAAAAGAACAGCAGGACCTTGATGAGAAAACCTGC 404

Qy 575 CAATTTATTCAGAGCAAGTGAAGAGGCTGTGAGGAGGAAAGAAAGAGGAGTCCCTAC 634
Db 405 CAATTTATTCAGAGCAAGTGAAGAGGCTGTGAGGAGGAAAGAAAGAGGAGGAGTCCCTAC 464

Qy 635 TTTTACGGAATTAAGCAGAGAAAATGATGAAGAGAAAGTCACTTTAATTTAGTAAAGG 694
Db 465 TTTTACGGAATTAAGCAGAGAAAATGATGAAGAGAAAGTCACTTTAATTTAGTAAAGG 524

Qy 695 AGCATGTAGCTATCCCGAGCAACATCTTCCAAAGTCAAGTACTCTGGACCGAGTGCAC 754
Db 525 AGCATGTAGCTATCCCGAGCAACATCTTCCAAAGTCAAGTACTCTGGACCGAGTGCAC 584

Qy 755 GAAGCAGATAGAGTTCAGCATCAGTGAAACGAAAGAAATCTTCCAGAGCTCAACTCA 814
Db 585 GAAGCAATATGAAAGTTGAGATCAGTGAAACGAAAGAAATCTTCCAGAGCTCAACTCA 644

Qy 815 GTCTAAAGAAAGAGAAAGAAATCTGCACTGTGATGAAATCATGGAGATTTGAAGAGGA 874
Db 645 GTCTAAAGAAAGAGAAAGAAATCTGCACTGTGATGAAATCATGGAGATTTGAAGAGGA 704

Qy 875 AAAGAAAAGAACTGCCGAAACAGACTACTGGCTACAGCTGAATTTATTTGAAAATTTAT 934
Db 705 AAAGAAAAGAACTGCCGAAACAGACTACTGGCTACAGCTGAATTTATTTGAAAATTTAT 764

Qy 935 AACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCTATTTGAAGGAGTAAATGA 994
Db 765 AACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCTATTTGAAGGAGTAAATGA 824

Qy 995 CAATATATACAGCTGTGTGAGATGATTTCTGGAGACAGCTGAAACTTTGACCAGAC 1054
Db 825 CAATATATACAGCTGTGTGAGAAATGATTTCTGAGAC-AGCTGAAACTTTGACCAGAC 883

Qy 1055 TCATTTAGAGCAGTAA-TTCCAGACCCAGGAAAAGAAATTTCTAGTTTT 1102
Db 884 TCATTTAGAGCAGTAA-TTCCAGACCCAGGAAAAGAAATTTCCAGTTTT 932

```

RESULT 15
 LOCUS CD109094
 DEFINITION AGENCOURT 14019646 NIH_MGC_179 Homo sapiens cDNA clone
 IMAGE:30366888 5', mRNA sequence.
 ACCESSION CD109094
 VERSION CD109094.1
 KEYWORDS GI:307622268
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.
 REFERENCE 1 (bases 1 to 890)
 AUTHORS NTH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rgs@nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 cDNA Library Preparation: Invitrogen Corp
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: NDAM427 row: d column: 01
 High quality sequence stop: 653.
 FEATURES
 source
 1. 890
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30366888"
 /tissue_type="Pituitary"
 /lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
 /clone_lib="NIH_MGC_179"
 /note="Organ: brain; Vector: pCMV-SPORT6.1; Site 1: EcoRV
 (destroyed); Site 2: NotI; Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.1 kb. Library was
 constructed by (Invitrogen). Note: this is a NIH_MGC
 Library."
 ORIGIN
 Query Match 48.6%; Score 742.6; DB 6; Length 890;
 Best Local Similarity 88.7%; Pred. No. 3.4e-159;
 Matches 867; Conservative 0; Mismatches 14; Indels 96; Gaps 2;
 Qy 27 AGAAGTGGGGTCCAGAAAGTGCCTGCGCGCTGCGCATGGGGAAGTCCGATTTTCTT 86
 Db 9 AGTACTGGGGTCCAGAAAGTGCCTGCGCGCTGCGCATGGGGAAGTCCGATTTTCTT 68
 Qy 87 ACTCCCAAGGCTATCGCCAAACAGATCAAGTCCCAAGGGGCTGCAGAGCTACGCTGAT 146
 Db 69 ACTCCCAAGGCTATCGCCAAACAGATCAAGTCCCAAGGGGCTGCAGAGCTACGCTGAT 128
 Qy 147 TCCAGATGTCAGAGAGCAGTGCAGGACAGAGTGGCTTTAAGTGTCTTATGATGCC 206
 Db 129 TGCCAGATGTCAGAGAGCAGTGCAGGACAGAGTGGCTTTAAGTGTCTTATGATGCC 161
 Qy 207 GAATCTCATCAGAGACAACTATTGCTGCGCTTCAGAAAAATCTCTCAGAGTATTATGAT 266
 Db 162 ----- 161
 Qy 267 TTTTCAGAGGAATCCGAAATGACTTTCTAGAACTTCTCAGAGACGCTTTGGCACTAAA 326
 Db 162 -----GGAATCCGAAATGACTTTCTAGAACTTCTCAGAGAGCGCTTTGGCACTAAA 213
 Qy 327 AGGTCCCAACAACATTTGCTACAAAGATATACATCAGCCACCGAGAGCAGCATCCACATG 386
 Db 214 AGGTCCCAACAACATTTGCTACAAAGATATACATCAGCCACCGAGAGCAGCATCCACATG 273

Qy	387	AATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGTGGCTGGCAGAGAGGCTTG	446
Db	274	AATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAGTGGCTGGCAGAGAGGCTTG	333
Qy	447	TGCAAAAGTGGACGAGACACCAGGCTGGTATATTTCAGTACATAGACAGGGACCCAGAA	506
Db	334	TGCAAAAGTGGACGAGACACCAGGCTGGTATATTTCAGTACATAGACAGGGACCCAGAA	393
Qy	507	ACTATCCGCGGCAACTGGAACTGGAGAAAAGAAAAGCAGGACCTTGATGATGAAGAA	566
Db	394	ACTATCCGCGGCAACTGGAACTGGAGAAAAGAAAAGCAGGACCTTGATGATGAAGAA	453
Qy	567	AAACTGCCCAATTTATTGAAGACCAAGTGAAGAGGCTGGAAGGGAAGGAACAGGAG	626
Db	454	AAACTGCCCAATTTATTGAAGACCAAGTGAAGAGGCTGGAAGGGAAGGAACAGGAG	513
Qy	627	GTCCCTACTTTTACGGAATTTAAGCAGAGAAAATGATGAAGAGAAAGTCAAGTTTAAATTTG	686
Db	514	GTCCCTACTTTTACGGAATTTAAGCAGAGAAAATGATGAAGAGAAAGTCAAGTTTAAATTTG	573
Qy	687	AGTAAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCAAGTCAAAGTCAAGTACTCTGGGACCG	746
Db	574	AGTAAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCAAGTCAAAGTCAAGTACTCTGGGACCG	633
Qy	747	AGTGCACCTGAAGACGATAGGAAGTTCAAGCATCAGTGAACGAAAAGAAATCTTCCAGAGC	806
Db	634	AGTGCACCTGAAGACGATAGGAAGTTCAAGCATCAGTGAACGAAAAGAAATCTTCCAGAGC	693
Qy	807	TCAACTCAGTCTAAAGAAAAGAAAAGAAATCTGCCTGGATGAAATCATGGAGATT	866
Db	694	TCAACTCAGTCTAAAGAAAAGAAAAGAAATCTGCCTGGATGAAATCATGGAGATT	753
Qy	867	GAAGAGGAAAAGAAAAGAACTGCCGCAACAGACTACTGGCTACAGCC-TGAAAATTATTGT	925
Db	754	GAAGAGGAAAAGAAAAGAACTGCCGCAACAGACTACTGGCTACAGCC-TGAAAATTATTGT	813
Qy	926	GAATAATTATACCAAGAACTGGAGAGAAATATCATAGAAAAGGCTATTGTTAAGGA	985
Db	814	GAATAATTATACCAAGAACTGGAGAGAAATATCTCTAAGAAAAGGCTATTGTTAAGGA	873
Qy	986	AGTAATTGACAAATATA 1002	
Db	874	AGTATTGACAAATATA 890	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:02 ; Search time 316.435 Seconds
(without alignment)
8583.479 Million cell updates/sec

Title: US-09-555-529-1

Perfect score: 1528
Sequence: 1 ctgaattcagggcgctg.....atgtattataaaaaaaaaa 1528

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/H COMB.seq.*
7: /cgn2_6/ptodata/1/ina/pp COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	151.4	9.9	23645	3	US-09-949-016-13916
C 2	123.6	8.1	538	3	US-09-270-767-379
C 3	123.6	8.1	538	3	US-09-270-767-15661
C 4	81	5.3	825	3	US-09-248-796A-5938
C 5	79.6	5.2	7218	2	US-08-232-463-14
C 6	53.2	3.5	25951	3	US-09-949-016-16194
C 7	51	3.3	1141	3	US-09-806-708B-22
C 8	50.6	3.3	1141	3	US-09-806-708B-22
C 9	46.2	3.0	147382	3	US-09-949-016-14624
C 10	45.8	3.0	209210	3	US-09-949-016-15094
C 11	45.4	3.0	612	3	US-09-902-540-1357
C 12	45.4	3.0	5181	2	US-08-257-073-10
C 13	44.6	2.9	1850	3	US-08-617-860B-32
C 14	44.6	2.9	4098	2	US-08-605-106-4
C 15	44.6	2.9	137394	3	US-09-949-016-13872
C 16	44.6	2.9	137743	3	US-09-949-016-12178
C 17	44.6	2.9	640681	3	US-09-790-988-1
C 18	44.4	2.9	37335	3	US-09-949-016-17132
C 19	44	2.9	4920	3	US-09-269-874A-1
C 20	43.6	2.9	114139	3	US-09-949-016-16536
C 21	43	2.8	612	3	US-09-902-540-1357
C 22	42.6	2.8	227750	3	US-09-949-016-17175
C 23	42.2	2.8	168174	3	US-10-071-411A-63
C 24	42.2	2.8	168273	3	US-10-071-411A-2

C 25	41.8	2.7	6243	2	US-09-056-075-1	Sequence 1, Appli
C 26	41.6	2.7	2181	3	US-09-662-254B-40	Sequence 40, Appli
C 27	41.6	2.7	2518	3	US-09-433-699-3	Sequence 3, Appli
C 28	41.6	2.7	2527	3	US-09-949-016-4169	Sequence 4169, Ap
C 29	41.6	2.7	2637	3	US-10-104-047-681	Sequence 681, App
C 30	41.6	2.7	50000	3	US-09-662-254B-24	Sequence 24, Appli
C 31	41.6	2.7	91062	3	US-09-949-016-13019	Sequence 13019, A
C 32	41.4	2.7	202001	3	US-09-734-674-3	Sequence 3, Appli
C 33	41.4	2.7	202001	3	US-10-274-990-3	Sequence 89871, A
C 34	41.2	2.7	601	3	US-09-949-016-89871	Sequence 93431, A
C 35	41.2	2.7	601	3	US-09-949-016-93431	Sequence 162984, A
C 36	41.2	2.7	601	3	US-09-949-016-162984	Sequence 14429, A
C 37	41.2	2.7	96987	3	US-09-949-016-14429	Sequence 43, Appli
C 38	41	2.7	2304	3	US-09-662-254B-43	Sequence 24, Appli
C 39	41	2.7	50000	3	US-09-662-254B-24	Sequence 3, Appli
C 40	40.8	2.7	3867	2	US-07-876-280-3	Sequence 3, Appli
C 41	40.8	2.7	3867	2	US-07-675-772-3	Sequence 3, Appli
C 42	40.8	2.7	3867	2	US-08-063-170-3	Sequence 3, Appli
C 43	40.8	2.7	3867	2	US-08-158-232-3	Sequence 3, Appli
C 44	40.8	2.7	3867	2	US-08-304-626-3	Sequence 3, Appli
C 45	40.8	2.7	3867	2	US-08-316-301A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-949-016-13916/c
; Sequence 13916, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13916
; LENGTH: 23645
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13916

Query Match	9.9%	Score 151.4; DB 3;	Length 23645;
Best Local Similarity	99.3%	Pred. No. 7.2e-28;	
Matches 152;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;
QY	27	AGAACTGGGGTCCAGAAAGTATCGCTCGGTGGTCCGATCGGGAAGTCGGATTTCTT	86
DB	1802	AGTACTGGGGTCCAGAAAGTATCGCTCGGTGGTCCGATCGGGAAGTCGGATTTCTT	1743
QY	87	ACTCCCAAGGGTATCGCCCAACAGGATCAAGTCCAGGGGCTCCAGAGCTACGCTGGTAT	146
DB	1742	ACTCCCAAGGGTATCGCCCAACAGGATCAAGTCCAGGGGCTCCAGAGCTACGCTGGTAT	1683
QY	147	TGCCAGATGTCACAGACAGTCGCCGGGACGAG	179
DB	1682	TGCCAGATGTCACAGACAGTCGCCGGGACGAG	1650

RESULT 2
US-09-270-767-379/c
; Sequence 379, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:


```
||||| |||||
Db 20110 ATATATTATAATA 20097

RESULT 7
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US/09/806,708B
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22
Query Match 3.3%; Score 51; DB 3; Length 1141;
Best Local Similarity 10.5%; Pred. No. 0.006;
Matches 110; Conservative 382; Mismatches 555; Indels 4; Gaps 2;

Qy 329 GGTCCACACACATTTGTCTACGAATACATACAGCCAGCAGACACATCCACATGAA 388
Db 1109 GKDNRMDATKWSATGTAWTTHAKRGATCMCYWYWTGTGTRRRCWRTYAMRTWYTRSNAN 1050
Qy 389 TGCACACTCAGTGGGAACTCTGACTGATTTACTAAGTGGCTGGCAGAGAGGCTTGTG 448
Db 1049 WSCATGKBMWMTKMYATKYRTAWMCAWRNNMWCATNGYAKSCATNNAMWYATTRWA 990
Qy 449 CAAGTGCAGCAGACACCAAAAGGCTGTATATTACGTACATAG---ACAGGGACCCAGA 505
Db 989 AYAAKAWARWAGNNRMRYGAAAGNKWGCMAAMATMBGWATAGKMCNNNNNNNTDVR 930
Qy 506 AACTATCCGCGGCAACTGGAACCTGAGAGAAAGAAAGAGACAGCCTTGATGATGAAGA 565
Db 929 MAMKAKNNNNNAYWTACYNRAATNNKMATMMKMTWTHGAHSKRTRHTRTCRTKYNN 870
Qy 566 AAAAAGCTCCCAATTTATTGAAGACCAAGTGAAGAGCGCTGGAGGGAAGGACAGGA 625
Db 869 NNNARTVYVYHHAARRMMNATWTRTNNNNNNNNNACRNTRTWABWKHSWCNNNNNN 810
Qy 626 GGTCCCTACTTTTACGGAATTAAGCAGAGAAATGATGAAGAGAAAGTCAAGTTTAATT 685
Db 809 NNNNNNTWCHYTTANABCBYRANNNNAARWARTCNMYHAAVTTTHDWCYKWTWNTWY 750
Qy 686 GAGTAAAGGACATGTAGCTCATCCGAGCAGCAACATCTTCCCAAGTCAAGTACTCTGGACC 745
Db 749 WDMTMTTBTTRNWT-TSTNMTNNNNNMWACTNNNNNNNNNNNNNNNNNNNNNNNNNNNN 691
Qy 746 GAGTGCATCAGACGATAGGAGTTCAGCATCAGTGAAGCAAGAAAGATCTTCCAGAG 805
Db 690 DARRNTNTVNRNRMTWTKTRWSTTRRHHTYGATNNNNNNNNNNNNNNNNNNNNNNNNNN 631
Qy 806 CTCAACTCAGCTAAAGAAAAGAAAAGAAATCTGCATCGATGAGTGAATCATGGAGAT 865
Db 630 MTMTWTKGDGMYTRKKYKWRDCTTYVDVWADSWWYVYANWRCRDVYTRNTYCKSY 571
Qy 866 TGAAGAGAGAAAGAAAGAACTGCCCGCAACAGACTACTGGCTACAGCCTGAAATTTGTT 925
Db 570 AHSYVYSSNNAMWYRRYSARNWSSMARWTTTRNNNNNNNNNNNNNNNNNNNNNNNNNN 511
Qy 926 GAAATATTATACCAAGAAACTGGGAGAGAAATATATCAAGAAAAGGCTATTGTTAAGGA 985

Db 510 RYYWKKRWABBTTTYVDSMCNAKSMWRGNNGNMMRAMKMMWAAANNNDAGAMDHTYMGNNMT 451
Qy 986 AGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTCTGGAGACAAGCTGAAACT 1045
Db 450 WMRRAKWMNMWACRRAYCCNNNNNRACVWHKHWRWTKYMWKAAACNNNNNBKAMYMRV 391
Qy 1046 TGACGAGACTCATTTAGAGACAGTAATTCAGCACCCAGGAAAAAGAAATCTTAGTTTAAA 1105
Db 390 AMMYSDTTNTDMMWMTSDWBWHYTVDTYMMRAMNNNNNNNNNNNNNNNNNNNNNNNNNN 331
Qy 1106 TGGAGGCTACAGAGGAATGAAGGTACCCCTAGATCCATCAATCAGAGAACTTTTTCAGC 1165
Db 330 THCTYGNNTWGSAYBMAASMSMAAGASNBVTYVNCWMTYMGKTMNTNNNNNNKAWYRTK 271
Qy 1166 TACTATCGTCAATTGAAACTGGCCCTTTTAAAGAGCAGCAGAGTTCGAAGAAATTCATATGA 1225
Db 270 TVAWCNRRYYDTAVWBTBKENYKYCYAYBMYBYMKGKHHBWMRRABHRSNNNNWVKCR 211
Qy 1226 AGACATTTTAAACTTCGCTGAGTTTGAAATTTGTTTAAACAATACATATAAACTTAAAG 1285
Db 210 NKYMVSWHYHAMRYBKWABAVGCNNNNNNKDRMAHHWCATNNNNNNNNNNNNNNNNNN 151
Qy 1286 CATCAAAATTCGTTTCGCCAAGGCATTTATGAGACTCTACTGTTGTTAGGCTATATCTTT 1345
Db 150 AWNNKKTABRDDHBAHVTKTYWRYDYWCAMCWNNAKAKVKTAMKHMWYTDTRYVSANNT 91
Qy 1346 GTATAAAACAAACAGGTTTTTTGAAATATTA 1376
Db 90 GVRMMWMMWCMWYVSMNNRMYRMRKYTWA 60

RESULT 8
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US/09/806,708B
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22
Query Match 3.3%; Score 50.6; DB 3; Length 1141;
Best Local Similarity 10.6%; Pred. No. 0.0076;
Matches 62; Conservative 243; Mismatches 279; Indels 2; Gaps 1;

Qy 943 AACTGGAGAGAAATATCATAAAGAAAAGGCTATTGTTAAGGAAGTAAATTGACAAATATA 1002
Db 91 ANNTSBRYHARRWKKDKMTAYBMTNKNWGTGRWRYRWRWMBDVTVDHHTVYVTAMNNAWT 150
Qy 1003 CAGCTGTGTGAAGATGATTGATTCTGGAGACAAGCTGAAACTTGCACGACTCATTTAG 1062
Db 151 TMCMDKDDKRETRWKKNNNNATGWDDDTKYHMMNNNGCBETVIMVRYKTRDWSBKRNN 210
Qy 1063 AGACAGTAATTCAGCACACCAGGAAAAAGAAATCTTAGTTTTTAAATGGAGGCTACAGAGAA 1122
Db 211 YGMBWKNWSYDVTYYVWVWDDMKCRKVRWRWTRGRMRNMYV--AMBTAAHRRYNNNGT 268
Qy 1123 ATGAGAGTACCTCAGATCCATCAATCAGAGAAAGCTTTTTCAGCTACTACTCGTCATTGAAA 1182
```

Db 269 BAAATRRWNNNNNNNAKCMKRAKXWGNBVBVNSTCTTWKSKTKTKVTSWANNCRAG 328
Qy 1183 CTGGCCCTTTAAAGAGCAGCAGAGTGAAGAAATCAATATAGAGACATTTCTAAACATG 1242
Db 329 DANKDHKKWKAAGVYNNNNNNNNNTYKARHBBARWVWHSAAKWKHANAHAHYSRKK 388
Qy 1243 CTGAGTTTGAATAATTTGTTAAACAATACATTAATTAATCTTAAGCATCAAAATGGTGTCG 1302
Db 389 WTBKRYKTMVNNNGTTTMMKMWYKMDMBWBTNNNNNGRTYYGWTNKKKKWITY 448
Qy 1303 CCAAGGCATTATGAGACTCTACTGTGTAGGGTATATCTTTTGTATATAAAACAAACAGGT 1362
Db 449 KWKANCKWRADPKTCHNTTWWKMTYNNCWKSWTKNGSHRBAAVYTWMMWR 508
Qy 1363 TTTTGAATAATATCTGTATAGTGTTCAGCTAAACTTTGAGAAGAAATTAATATGCTT 1422
Db 509 RVAHANNNNDYWKACTYKVBVCCKWNNYAAWYTKSSMNYTSRYVRWKTNNSWRWS 568
Qy 1423 CATGAGTATCAAACTATGTAATTTTGTCTTGTATTTTCTTCTTCTTGTAAATTTACT 1482
Db 569 DTRSMGRANNYARABHYGYKWNTRWBSHTWBHBRAGAHHYMBWYBAKCHCKWATK 628
Qy 1483 TGATGAGTTTATATCTTTCATTAAAGAACTGTATTATATAAAAAA 1528
Db 629 AKYAGAGSNNNNNNNNNNNNNNNNATCARDYYAASRWYMANA 674

RESULT 9
US-09-949-016-14624
; Sequence 14624, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14624
; LENGTH: 147382
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(147382)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14624

Query Match 3.0%; Score 46.2; DB 3; Length 147382;
Best Local Similarity 52.3%; Pred. No. 0.72;
Matches 102; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
Qy 1333 GGTATATCTTTGTTAAACAAACAGGTTTGAATAATATCTGTATAGTTGTCAG 1392
Db 136880 GTTATATATTTTATATATAGTTATATATATATATATATATATATATATATATAT 136939
Qy 1393 CTAACTTCGAGAGATTTAATATGTCATGAGTATCAAACTATGAAATTTGTCC 1452
Db 136940 ATATATGATATATATTAACATTTTATATATATATATATATATATATATATATAT 136999
Qy 1453 TTGTTATTTTGTTCCTTGTAAATTTTACGTGAGTTTATATCTTCATTAAGAATGT 1512
Db 137000 TAATATATTTAT 137059

Qy 1513 TATTATAAAAAA 1527
Db 137060 ATATATATATATA 137074
RESULT 10
US-09-949-016-15094
; Sequence 15094, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15094
; LENGTH: 209210
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(209210)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15094

Query Match 3.0%; Score 45.8; DB 3; Length 209210;
Best Local Similarity 50.6%; Pred. No. 1;
Matches 136; Conservative 0; Mismatches 132; Indels 1; Gaps 1;
Qy 1251 TGAAAATTTGTTAAACAATACATTAATAATCTTAAGCATCAAAATGGTGTGCCAAGCA 1310
Db 112647 TGAAGTATCTTACATGTCAGTTAAATCTGATGGTTTATATATCTGTCAGTATTTC 112706
Qy 1311 TTATGAGACTCTACTGTGTAGGGTATATCTTTTGTATAAAACAAACAGGTTTGTAAA 1370
Db 112707 TTATTAATCTTCTGTATAGTTTGTCCATTTTGAAGTGAGATATTGAAGTCTCAAAA 112766
Qy 1371 ATATTACTGTATAGTTGTTCAGCTAAACTTTGAGAAGAAATTTAATATGTCCTCATGAGT 1430
Db 112767 CTAATATTTTGGAGTTGCTATTTCTCTCTTTTCAGTTTTCATATATTTTATATAT 112826
Qy 1431 ATCAAACTATGTAATTTGTCTCTCTCT-TATTTTGTCTTCTTGTAAATTTACTTGATGAG 1489
Db 112827 TTGGCACTCCGTTGTTAGTTCATGCTGTTTATTAATTTGTTATATCTTCTGTTGATG 112886
Qy 1490 TTTATATCTTCAATAAAGAAATGTTATTAT 1518
Db 112887 ATCCCTTTATCATAGAAAATGCTTTGT 112915

RESULT 11
US-09-902-540-1357
; Sequence 1357, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883

```
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1357
; LENGTH: 612
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(612)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1357

Query Match          3.0%; Score 45.4; DB 3; Length 612;
Best Local Similarity 43.3%; Pred. No. 0.12;
Matches 205; Conservative 0; Mismatches 268; Indels 0; Gaps 0;

QY 532 AGAAAAAGAAAAAGCAGGACCTGTGATGATGAAGAAAAAAGTCCCAAAATTTTGAAGAGC 591
DB 80 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 139
QY 592 AGTGAAGAGGCTGGAAGGGAAGGACAGGAGGTCCCTACTTTTACGGAATTAAGCA 651
DB 140 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 199
QY 652 GAGAAATCATGAAGAGAAAGTCACGTTTAAATTTGAGTAAGGAGCATGTGCTCATCG 711
DB 200 AATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 259
QY 712 GAGCAACATCTTCCAAAGTCAGTACTCTGGGACCGAGTGCACTGAAGACGATAGGAAGTT 771
DB 260 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 319
QY 772 CAGCATCAGTGAACGAAAGAAATCTTCCGAGCTCAACTCAGTCTAAAGAAAAAGAGA 831
DB 320 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 379
QY 832 AAAAGAAATCTGCATCGATGAAATCATGGAATTTGAAGAGGAAAGAAAGAAAGCTGCC 891
DB 380 AAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 439
QY 892 GAACAGACTACTGGCTACAGCCCTGAAATTTATTTGTGAAATTTATAACCAAGAACTGGAG 951
DB 440 TTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 499
QY 952 AGAATATCATAGAAAAAGGCTATTGTTAAGGAAGTAAATGACAAATATACA 1004
DB 500 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 552

RESULT 12
US-08-257-073-10/c
; Sequence 10, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
```

```
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Prommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-0712
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5181 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-257-073-10

Query Match          3.0%; Score 45.4; DB 2; Length 5181;
Best Local Similarity 51.1%; Pred. No. 0.29;
Matches 158; Conservative 0; Mismatches 146; Indels 5; Gaps 2;

QY 1223 TGAAGACATTTCTAAATTCGCTGAGTTTGAAATTTGTTAAACAATACATATAAATCTTA 1282
DB 3483 TAAAGTTTAAATGGAGATGATTCACCATTAATAATATTAACAAGTCTCTTATAATGTTT 3424
QY 1283 AAGCATCAAAATTTGTTTCGCCAAGGCATTAATGAGACTCTACTGTTAGGATATATTTCT 1342
DB 3423 CAATAATATTTTGTGTTTCTAATGTTTTCAGTTTCGTCTATTTTCAGCTCTCTTTT 3364
QY 1343 TTTGT---ATAAACAAACAGGTTTTTTGAAAAATATTAAT---GTATAGTTGTTTCAGCTAAA 1397
DB 3363 TTTCTTAAAGAAACACAGAAAAGTTTGTATAATACATATGTTGGTTATTAAGTGAATTCAA 3304
QY 1398 CTTTGAAGAAGATTTAATTAATGCTCATGAGGTATCAAACTATGTAATTTTGCTGTTT 1457
DB 3303 TTTTGATTTCAATTTGTTCTTTTAAATAAAGTAAGTTTTTAAATTTTGCATTTTGTCTGGCC 3244
QY 1458 ATTTTGTCTTCTTTGTAATTTACTTCATGAGTTTATATCTTCAATAAAGAAATGTTATTA 1517
DB 3243 AAGTTCTTTTCTTATTAATAATATCTAATTTTAATTTATATATATATAAATC 3184
QY 1518 TAAAAAAA 1526
DB 3183 AAATGATAA 3175

RESULT 13
US-08-617-860B-32
; Sequence 32, Application US/08617860B
; Patent No. 6133506
; GENERAL INFORMATION:
; APPLICANT: Tvpfer, R., Bautor, J., Bothmann, H., Filsak, E.,
; APPLICANT: Hvricke-Grandpierre, C., Klein, B., Martini, N.,
; APPLICANT: Miller, A., Schulte, W., Voetz, M., Walek, J.,
; APPLICANT: Schell, J.
; TITLE OF INVENTION: Promoters
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
; STREET: 1140 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
```


[illegible]

```
; FEATURE: intron IV
; NAME/KEY: intron IV
; LOCATION: 3012..3131
; FEATURE:
; NAME/KEY: exon V
; LOCATION: 3132..3303
; FEATURE:
; NAME/KEY: intron V
; LOCATION: 3304..3390
; FEATURE:
; NAME/KEY: exon VI
; LOCATION: 3391..3459
; FEATURE:
; NAME/KEY: intron VI
; LOCATION: 3460..3671
; FEATURE:
; NAME/KEY: exon VII
; LOCATION: 3672..3941
; FEATURE:
; NAME/KEY: Stopcodon
; LOCATION: 3942..3944
;
US-08-605-106-4

Query Match      2.9%; Score 44.6; DB 2; Length 4098;
Best Local Similarity 51.4%; Pred. No. 0.43;
Matches 128; Conservative 0; Mismatches 119; Indels 2; Gaps 1;

QY 1271 ATTAAATCTTAAAGCATCAAAATCGTGTCGCCAAGGCATTATGAGACTCTACTGCTGTT 1330
Db 141 ATTTAAATTTTGGCGGATATATTTGTAATATTTTATGAATTTATGAAATATTTTTCGAA 200

QY 1331 AGGGTATATCTTTTGTTATAAACAAACAGGTTTTTGAATAATTTACTGTATAGTTGTC 1390
Db 201 ATTTAAATATTTTAAATTTTAAATTTTAAATATATTTTAAATTTCTTTTAAAAAATAATTTT 260

QY 1391 AGCTAAACTTTGAGAGAAATTTAAATTTATCTCATGAGTATCAAACTATGTAATTTTGT 1450
Db 261 AAATATTATAAATTTAGTTTTTAAATTTTTTAA--ATATTTAAATTTAGTTTTTTTAA 318

QY 1451 CTTGTTATTTTGTTCCTTTGTTAAATTTACTGTAGTATTTATCTTCATTAAGAAAT 1510
Db 319 TTTTAAATATTTGTTGAATTTTAAATATATTTTGGTTTTTAAATATATATTTTAAAGT 378

QY 1511 GTTATTATA 1519
Db 379 TTTTAAATA 387
```

```
RESULT 15
US-09-949-016-13872/c
; Sequence 13872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13872
; LENGTH: 137394
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (1)....(137394)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13872

Query Match      2.9%; Score 44.6; DB 3; Length 137394;
Best Local Similarity 53.1%; Pred. No. 1.8;
Matches 95; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 820 AAGAAAAAGAGAAAAAGAAATCTGCACTGGATGAAATCATGAGATTGAAAGAGAAAAAGA 879
Db 15531 AAGAAAAAGAGAAAAAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAGAGAAAGA 15472

QY 880 AAAGAACTGCCCGAACACAGACTACTGGCTACAGCTGAAATTTTGTGAAAATTTATAACCA 939
Db 15471 AAGGAAAGAGAAAGGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAGAAAGAGAAAGAA 15412

QY 940 AGAAACTGGGAGAGAAAAATATCATAAGAAAAAGGCTATTGTTAAGGAAAGTAAATTCACAAA 998
Db 15411 AGAAAGAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAGAGAAAGAGAGAA 15353

Search completed: November 27, 2005, 01:10:36
Job time : 319.435 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 : Search time 1761.53 Seconds
(without alignments)
7173.088 Million cell updates/sec

Title: US-09-555-529-1
Perfect score: 1528
Sequence: 1 ctgaattcagcgccgctg.....atgtattataaaaaaaaa 1528

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA Main:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	490.8	32.1	591	5	US-10-106-698-1187 Sequence 1187, Ap
2	436.4	28.6	461	7	US-10-242-535A-29177 Sequence 29177, A
3	436.4	28.6	461	7	US-10-085-783A-29177 Sequence 29177, A
4	298.6	19.5	1241	10	US-11-097-143-41039 Sequence 41039, A
5	284	18.6	300	9	US-10-779-543-7646 Sequence 7646, Ap
6	247.2	16.2	3313	10	US-11-097-143-41038 Sequence 41038, A
7	209.4	13.7	4394	10	US-11-097-143-41017 Sequence 41017, A
8	209.4	13.7	4582	10	US-11-097-143-24532 Sequence 24532, A
9	205.6	13.5	777	8	US-10-424-599-25484 Sequence 25484, A
10	204.8	13.4	1713	8	US-10-739-930-2955 Sequence 2955, Ap
11	203.2	13.3	2046	8	US-10-425-115-157125 Sequence 157125, Ap
12	201.6	13.2	1614	7	US-10-425-114-2975 Sequence 2975, Ap
13	201.6	13.2	1713	7	US-10-425-114-22614 Sequence 22614, A
14	192	12.6	538	7	US-10-437-963-89433 Sequence 89433, A
15	156.8	10.3	584	7	US-10-767-701-25587 Sequence 25587, A
16	126.4	8.3	549	3	US-09-991-936-1424 Sequence 1424, Ap
17	126.4	8.3	549	9	US-10-978-245-1424 Sequence 1424, Ap
18	116	7.6	116	7	US-10-242-535A-7473 Sequence 7473, Ap
19	116	7.6	116	7	US-10-085-783A-7473 Sequence 7473, Ap
20	106.8	7.0	739	7	US-10-424-599-53324 Sequence 53324, A
21	102.4	6.7	516	4	US-09-925-065A-549989 Sequence 549989, Ap
22	100.2	6.6	270	3	US-09-294-093B-3089 Sequence 3089, Ap
23	78.4	5.1	588	4	US-09-925-065A-20834 Sequence 20834, A

24	72.6	4.8	1608	4	US-09-925-065A-707309 Sequence 707309, A
25	71.2	4.7	616	7	US-10-424-599-53326 Sequence 53326, A
26	60	3.9	60	3	US-09-908-975-6340 Sequence 6340, Ap
27	59.8	3.9	521	8	US-10-425-115-116286 Sequence 116286, Ap
28	59.4	3.9	12578	7	US-10-221-714A-382 Sequence 382, App
29	58	3.8	5935	5	US-10-239-676-134 Sequence 134, App
30	56.8	3.7	2985	10	US-11-097-143-41056 Sequence 41056, A
31	56	3.7	3673778	6	US-10-312-841-1 Sequence 1, Appli
32	55.2	3.6	358	7	US-10-424-599-106049 Sequence 106049, A
33	55.2	3.6	11422	6	US-10-311-455-192 Sequence 192, App
34	55.2	3.6	11422	7	US-10-257-166-18 Sequence 18, Appl
35	54.4	3.6	65	3	US-09-908-975-24642 Sequence 24642, A
36	54	3.5	15951	6	US-10-311-455-1653 Sequence 1653, Ap
37	54	3.5	15951	6	US-10-240-485-133 Sequence 133, App
38	53.8	3.5	1204	7	US-10-437-963-77858 Sequence 77858, A
39	53.8	3.5	6831	6	US-10-311-455-1459 Sequence 1459, App
40	53.4	3.5	6013	7	US-10-221-613-225 Sequence 225, App
41	53	3.5	8056	8	US-10-473-126-386 Sequence 386, App
42	52.6	3.4	113515	6	US-10-311-455-2148 Sequence 2148, Ap
43	51.8	3.4	3720	5	US-10-172-086-74 Sequence 74, Appl
44	51.8	3.4	3720	7	US-10-311-507-38 Sequence 38, Appl
45	51.8	3.4	3720	8	US-10-480-846-74 Sequence 74, Appl

ALIGNMENTS

RESULT 1

US-10-106-698-1187
; Sequence 1187, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005PI
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1187
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (587)..(587)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1187

Query Match 32.1%; Score 490.8; DB 5; Length 591;

Best Local Similarity 97.0%; Pred. No. 1.6e-109;
Matches 512; Conservative 0; Mismatches 12; Indels 4; Gaps 1;

QY	1005	GCTGTTGTGAAGATGATTTCTCGAGACAAGCTGAAACTTGACCAGACTCATTTAGAG	1064
Db	28	GCTGCAGGAATTCGGCAGCAGTCTGGAGACAAGCTGAAACTTGACCAGACTCATTTAGAG	87
QY	1065	ACAGTAATTCAGCACCAGGAAAAAGAAATTCAGTTTAAATGGAGGCTACAGAGAAAT	1124
Db	88	ACAGTAATTCAGCACCAGGAAAAAGAAATTCAGTTTAAATGGAGGCTACAGAGAAAT	147
QY	1125	GAAGGTACCTAGAAATCCATCAATCAGAGAGACTTTTTCAGTACTATCGTCTATTGAACT	1184
Db	148	GAAGGTACCTAGAAATCCATCAATCAGAGAGACTTTTTCAGTACTATCGTCTATTGAACT	207
QY	1185	GGCCCTTTTAAAGGACGACAGTTGAAGGAATTCATATGAAGACATTTCTAACTTGGC	1244

```
Db 208 GCGCCCTTTAAAGGACGACAGTTCGAGGAATTCATATGAGACATTTCTAAACTTGCC 267
QY 1245 TGAGTTTGAAATTTGTTAAACAATCATTAATAATCTTAAAGCATCAAAATGGTGTCCGC 1304
Db 268 TGAGTTTGAAATTTGTTAAACAATCATTAATAATCTTAAAGCATCAAAATGGTGTCCGC 327
QY 1305 AAGCATTATGAGACTCTACTGTGTAGGTTATATCTTTTGTATATAAACAACAGGTTT 1364
Db 328 AAGCATTATGAGACTCTACTGTGTAGGTTATATCTTTTGTATATAAACAACAGGTTT 387
QY 1365 TTGAAATATTAATCTGTA----TAGTTGTTTACGCTAAACTTTTGAGAGAATTTAAATATGT 1420
Db 388 TTGAAATATTAATCTGTAATAGTTAGTTGTTTACGCTAAACTTTTGAGAGAATTTAAATATGT 447
QY 1421 CTCATGAGGTATCAAACTATGTAATTTTGTCTCTGTTATTTTGTCTTTGTTGTAATTTA 1480
Db 448 CTCATGAGGTATCAAACTATGTAATTTTGTCTCTGTTATTTTGTCTTTGTTGTAATTTA 507
QY 1481 CTTGATGAGTTTATATCTTCAATTAAGAATGTTTATTATATAAAAAAAA 1528
Db 508 CTTGATGAGTTTATATCTTCAATTAAGAATGTTTATTATATAAAAAAAA 555

RESULT 2
US-10-242-535A-29177
; Sequence 29177, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29177
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-29177

Query Match 28.6%; Score 436.4; DB 7; Length 461;
Best Local Similarity 99.4%; Pred. No. 2.9e-96;
Matches 459; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 59 GGTCCCATGGGAAGTCGGATTTCTTACTCCCAAGCTATCGCCACAGGATCAAGTC 118
Db 1 GGTCCCATGGGAAGTCGGATTTCTTACTCCCAA-GCTATCGCCAAACAGGATCAAGTC 59
QY 119 CAAGGGCTGCAGAAAGCTACGCTGTTATTCAGAGATGTCAGAAAGCAGTCCCGGACGA 178
Db 60 CAAGGGCTGCAGAAAGCTACGCTGTTATTCAGAGATGTCAGAAAGCAGTCCCGGACGA 119
QY 179 GAATGGCTTTAAGTGTCAATTTGATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTC 238
Db 120 GAATGGCTTTAAGTGTCAATTTGATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTC 179
QY 239 AGAAATCCTCAGCAGTTTATGGAATATTTTTCAGAGGAATTCGAAATGACTTTCTAGA 298
Db 180 AGAAATCCTCAGCAGTTTATGGAATATTTTTCAGAGGAATTCGAAATGACTTTCTAGA 239
QY 299 ACTTCTCAGGAGACGCTTTTGGCACTAAAGGGTCCCAACAACATTTCTTACACGAATA 358
Db 240 ACTTCTCAGGAGACGCTTTTGGCACTAAAGGGTCCCAACAACATTTCTTACACGAATA 299
QY 359 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 418
Db 300 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 359
QY 419 TACTTAAGTGTCTGGC-AGAGAAGGCTTGTGCAAAAGTGGACGAGACACCAAAAGGCTGGT 477
Db 360 TACTTAAGTGTCTGGCAGAGAGAGAGGCTTGTGCAAAAGTGGACGAGACACCAAAAGGCTGGT 419
QY 478 ATATTCAAGTACATAGACAGGACCCAGAAACTATCCGCGGC 519
Db 478 ATATTCAAGTACATAGACAGGACCCAGAAACTATCCGCGGC 519
```

```
QY 359 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 418
Db 300 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 359
QY 419 TACTTAAGTGTCTGGC-AGAGAAGGCTTGTGCAAAAGTGGACGAGACACCAAAAGGCTGGT 477
Db 360 TACTTAAGTGTCTGGCAGAGAGAGAGGCTTGTGCAAAAGTGGACGAGACACCAAAAGGCTGGT 419
QY 478 ATATTCAAGTACATAGACAGGACCCAGAAACTATCCGCGGC 519
Db 420 ATATTCAAGTACATAGACAGGACCCAGAAACTATCCGCGGC 461

RESULT 3
US-10-085-783A-29177
; Sequence 29177, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29177
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-29177

Query Match 28.6%; Score 436.4; DB 7; Length 461;
Best Local Similarity 99.4%; Pred. No. 2.9e-96;
Matches 459; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

QY 59 GGTCCCATGGGAAGTCGGATTTCTTACTCCCAAGCTATCGCCACAGGATCAAGTC 118
Db 1 GGTCCCATGGGAAGTCGGATTTCTTACTCCCAA-GCTATCGCCAAACAGGATCAAGTC 59
QY 119 CAAGGGCTGCAGAAAGCTACGCTGTTATTCAGAGATGTCAGAAAGCAGTCCCGGACGA 178
Db 60 CAAGGGCTGCAGAAAGCTACGCTGTTATTCAGAGATGTCAGAAAGCAGTCCCGGACGA 119
QY 179 GAATGGCTTTAAGTGTCAATTTGATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTC 238
Db 120 GAATGGCTTTAAGTGTCAATTTGATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTC 179
QY 239 AGAAATCCTCAGCAGTTTATGGAATATTTTTCAGAGGAATTCGAAATGACTTTCTAGA 298
Db 180 AGAAATCCTCAGCAGTTTATGGAATATTTTTCAGAGGAATTCGAAATGACTTTCTAGA 239
QY 299 ACTTCTCAGGAGACGCTTTTGGCACTAAAGGGTCCCAACAACATTTCTTACACGAATA 358
Db 240 ACTTCTCAGGAGACGCTTTTGGCACTAAAGGGTCCCAACAACATTTCTTACACGAATA 299
QY 359 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 418
Db 300 CATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 359
QY 419 TACTTAAGTGTCTGGC-AGAGAAGGCTTGTGCAAAAGTGGACGAGACACCAAAAGGCTGGT 477
Db 360 TACTTAAGTGTCTGGCAGAGAGAGAGGCTTGTGCAAAAGTGGACGAGACACCAAAAGGCTGGT 419
QY 478 ATATTCAAGTACATAGACAGGACCCAGAAACTATCCGCGGC 519
Db 478 ATATTCAAGTACATAGACAGGACCCAGAAACTATCCGCGGC 519
```

Db 420 ATATTAGTACATAGACAGGAGCCAGAACTATCCGCCGC 461

RESULT 4

US-11-097-143-41039
; Sequence 41039, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41039
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41039

Query Match 19.5%; Score 298.6; DB 10; Length 1241;
Best Local Similarity 54.2%; Pred. No. 3e-62;
Matches 638; Conservative 0; Mismatches 524; Indels 15; Gaps 1;

QY 64 CCATGGGAGTCGGATTTCTTACTCCCAAGCTATCGCCAAACAGATCAAGTCCAAG 123
DB 67 CGATGGTCGCGCGAGGTAGGTACGCCCAAGTACCTCGCCAAACAGATCAAGTCCAAG 126

QY 124 GGCTGCAGAGCTACGCTGGTATGCCAGATGTCCAGAGCAGTCCCGGACGAGATG 103
DB 127 GTCTGCAGAGCTGCGCTGGTACTGCCAGATGTCCAGAGCAGTCCCGGATGAACAG 186

QY 184 GCTTTAAGTGTCAATTGTATGTCGCAATCTCATCAGAGACAACTATTGCTGGCTTCAGAA 243
DB 187 GCTTCAAGTCCACAGATGAGCGAGTCCACAGCGCCAGTGTCTCTCTTTCGGGACA 246

QY 244 ATCCTCAGCAGTTTATGGATTTATTTTTCAGAGGAATTCGAAATGACTTTCTAGACTTC 303
DB 247 ATCCTGCAATTTCTGACAGCTTCAGCAAGAGTTCCTCGACGGCTACATGGAGTTGC 306

QY 304 TCAGGAGACGCTTGGCACTAAAGGGTCCCAACACATGTTCTACACGAATATACATCA 363
DB 307 TGGCGCGCGGTTTCGGCACCAAGCGAACACAGCTCTACAGAGTACATG 366

QY 364 GCCACCGAGAGCAGATCCATGAATGCTCACTAGTGGGAACTCTGACTGATTTTACTA 423
DB 367 CCACAAAGGAGCAGATCCACATGACGCCACCGGATGGCTCACCTGCTCGACTACGTGA 426

QY 424 AGTGGCTGGGCGAGAGAGGCTTGTGCAAGTGGACGACACCAAAAGGCTGGTATATTC 483
DB 427 AGTGGCTGGGCGGAGCTGGGCAAGTGTAGCGATGAGACGGAGAGGCTGGTTCGTCA 486

QY 484 AGTACATAGACAGGAGCCAGAAACTATCCCGCGGCAACTGGAACCTGGAGAAAAGAAA 543

Db 487 CCTACATTGATCGCAGTCCAGAGGCCATGGAACGACAGCGGAGGTGATCGCAAGGAGA 546
QY 544 AGCAGGACCTTGATGATGAAGAAAAATGCGCAAAATTTTTCAGAGGCAAGTGCAGAGAG 603
DB 547 AGATGGAGAAAGGACGACGAGGAGCGGATGCGCACTTTCATTGAGCAGCAGATTTAAATG 606
QY 604 GCCTGGAAGGGAAGCAACAGGAGGTCCCTACTTTTTCGGAATTAAGCAGAGAAAAATGATG 663
DB 607 CCAAGGCCNAGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 666
QY 664 AAGAGAAAGTCACTGTTTAAATTTTGAAGTAAAGGAGCATGTAGCTCATCCGAGGACAACTCTT 723
DB 667 AAGAGAAACGAACCACTCAAGCTTGATATTCG-----CCTTGAGAAAAAGT 711
QY 724 CCAAGTCAAGTACTCTGGGACCGAGTGCACCTGAAGCAGATAGAGAGTTCAGCATCATGTA 783
DB 712 TCCAGCCTGACACTGTGTAGGGAATCCGCTCTAGCAGGAGCATCGCCCTTGAAGCTG 771
QY 784 AACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAGAAAGAAATCTG 843
DB 772 AAGAAAGGTGTTCAAGAAACCAAAATCCGTGGCTGGAGACAGCCAAACCGCGTCGTCG 831
QY 844 CACTGATGAAATCATGGAGATTGAAGAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGT 903
DB 832 TGGACGAGATCATCAAGCAGGAGGAGGAAAGCAAAAGGAGGCTGCCAACCGCAGGACTACT 891
QY 904 GGCTACAGCTCAAAATTTTGAAGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 963
DB 892 GGCTGCACAAAGGATATCGTGGTCAAAATTTTTCGCAATTCATGGGCGGAAAGTCTTCA 951
QY 964 AGAAAGGCTATTGTTAAGGAAAGTAAATGACAAATATATACAGCTGTTGTAAGATGATG 1023
DB 952 AACAAAGCGGTTGCTCGGAGCTAATTTGACAGATATCAGGCGCAAAATCAAGTCTTGG 1011
QY 1024 ATTTCTGGAGACAAAGCTGAAACTTGACCAGACTATTTAGAGCAGATTAATTCAGACACAG 1083
DB 1012 AGACTGGGAAAGCTAAAGTGGATCAAGCTCATTTGGAGAGCGTAAATCCCGCTTTGG 1071
QY 1084 GAAAGAAATTTAGTTTAAATGGAGGCTACAGAGGAAATGAAGTACCTACCTAGATCCA 1143
DB 1072 ACAAGCTGTCTATGGTGGTAAATGGCGCTTATCGGGGATCCGAGGCTCTGCTAAGGAAAC 1131
QY 1144 TCAATCAGAGAGCTTTTTCAGCTACTATCGTCAATTAAGAACTGGCCCTTTAAAGGAGCGCA 1203
DB 1132 TGGACGAGGCGAGATATTGATCAGCGCTGGAATATTTGACGCTCTCTCAAGGCGAGAA 1191
QY 1204 GAGTTGAAGGAAATCAATATGAAGACATTTCTAAACT 1240
DB 1192 TTGTAGACAAAGCTGCAATACGAAGATATATCTAAACT 1228

RESULT 5
US-10-779-543-7646
; Sequence 7646, Application US/10779543
; Publication No. US20050227917A1
; GENERAL INFORMATION:
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
; TITLE OF INVENTION: IN CANCEROUS CELLS AND THEIR METHODS OF USE II
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/076,555
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 09/217,471
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: 60/068,755
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/080,664
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/105,234
; PRIOR FILING DATE: 1998-10-21

```
; PRIOR APPLICATION NUMBER: 09/297,648
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: PCT/US99/01619
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: 60/072,910
; PRIOR FILING DATE: 1998-01-28
; PRIOR APPLICATION NUMBER: 60/075,954
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/080,114
; PRIOR FILING DATE: 1998-03-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7646
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-779-543-7646

Query Match      18.6%; Score 284; DB 9; Length 300;
Best Local Similarity 10.0%; Pred. No. 5.1e-59;
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 634 CTTTACCGAATTAGCAGAGAAATGATGAGAGAAAGTCAAGTTTAAATTTGACTAAAG 693
Db 17 CTTTACCGAATTAGCAGAGAAATGATGAGAGAAAGTCAAGTTTAAATTTGAGTAAAG 76

QY 694 GAGCATGTAGCTCATCCGGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCAC 753
Db 77 GAGCATGTAGCTCATCCGGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCAC 136

QY 754 TGAAGACATAGGAAGTTTCAGCATCAGTGAACGAAACGAAAGAAATCTTCCAGAGCTCAACTC 813
Db 137 TGAAGACATAGGAAGTTTCAGCATCAGTGAACGAAACGAAAGAAATCTTCCAGAGCTCAACTC 196

QY 814 AGTCTAAAGAAAGAAAGAAAGAAATCTGCACTCGATGAAATCATGAGATTGAAGAGG 873
Db 197 AGTCTAAAGAAAGAAAGAAAGAAATCTGCACTCGATGAAATCATGAGATTGAAGAGG 256

QY 874 AAAAGAAAAGAACTGCCGAAACAGACTACTGGCTACAGCTGAA 917
Db 257 AAAAGAAAAGAACTGCCGAAACAGACTACTGGCTACAGCTGAA 300

RESULT 6
US-11-097-143-41038
; Sequence 41038, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; PRIOR FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23

; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41038
; LENGTH: 3313
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41038

Query Match      16.2%; Score 247.2; DB 10; Length 3313;
Best Local Similarity 54.1%; Pred. No. 2e-49;
Matches 535; Conservative 0; Mismatches 438; Indels 15; Gaps 1;

QY 64 CCATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAAGAGTCAAGTCCAAAG 123
Db 1067 CGATGGGTGCGCGCGAGGTAGGTACGCCCAAGTACTCGCCAAAGAGTGAATGAATCGAAGG 1126

QY 124 GGCTGCAGAAAGCTACGCTGGTATTGCCAGATGTGCCAAGCAGTCCCGGAGCAGGAATG 183
Db 1127 GTCTGCAGAAAGCTGGCTGGTACTGCCAGATGTGCCAAGCAGTCCCGGAGTGAATG 1186

QY 184 GCTTTAAGTGTCAATGTATGTCCGAAATCTCATACAGAGACAATATGTGTGGCTTCAGAAA 243
Db 1187 GCTTCAAGTGCACACGATGAGCGAGTCCCAACGAGCGCAGTTGCTCTCTTTGGGACA 1246

QY 244 ATCCTCAGCAGTTTATGGATTATTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTC 303
Db 1247 ATCCTGGCAAAATCCTGCACAGCTTCAGAAAGAGTTCTCCGACGGCTACATGGAGTTGC 1306

QY 304 TCAGAGAGACGCTTTGGCACTAAAAAGGTCACAAACAACTTGTCTTACAAAGAAATACATCA 363
Db 1307 TGCGCCGCGGTTTGGCAGCAGCGAACCCAGCGCCACAAAGATCTACCAGAGTACATTG 1366

QY 364 GCCACCGAGAGACATCCACATGAATGCCATCAGTGGGAAACTCTGTGACTGATTTTACTA 423
Db 1367 CCCACAAGGAGACATCCACATGAACGCCACCGAGTGTCTCACTGTCCGACTACGTGA 1426

QY 424 AGTGGCTGGGAGAGAGGCTTGTGCAAGTGGAGCAGACACCAAAAGGCTGGTATATTC 483
Db 1427 AGTGGCTGGGCGGAGTGGGCAAGTGTAGAGCGGATGAGACGGAGAGAGGCTGGTTCGTCA 1486

QY 484 AGTACATAGACAGGAGACCCAGAAACTATCCCGCGGCAACTGGAACTGGAGAGAAAAGAAA 543
Db 1487 CCTACATTGATCGCAGTCCAGAGGCCATGGAAACGACAGCGCGAGGCTGATCCGAAGGAG 1546

QY 544 AGCAGGACCTTGATGATGAAGAAAACCTGCCAAATTTATTGAAGAGCAAGTGAAGAAAG 603
Db 1547 AGATGGAGAGAGACGACGAGGAGCGGATGGCGGACTTCATTGAGCAGCAGAGATTAAAAATG 1606

QY 604 GCCTGGAAGGGAAGGACGAGGAGTCCCTACTTTTACGGAATTAAGCAGAGAAAATGATG 663
Db 1607 CCAAGGCCAAGGACGCGGAGGAGGAGCAGAGGCCAGGAGAGAGTTTACCAGCTAAAGCGCG 1666

QY 664 AAGAGAAAGTCACGTTTAAATTTGATGTAAGGAGCATGTAGCTCATCCGAGCAACATCTT 723
Db 1667 AAGAGAACGAACCACTCAAGCTTGTATTCG-----CCTTGAGAAAAAGT 1711

QY 724 CCAAGTCAAGTACTCTGGGACCGAGTGCACTGAAGACGATAGGAAAGTTTACGATCAGTGA 783
Db 1712 TCCAGCCTGACACTGTGTAGGGAAATCCGCTCTAGCCAAGCGACCTGCCCCCTGAAGCTG 1771

QY 784 AACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAAGAAAGAAAAGAAATCTG 843
Db 1772 AAGAAAAGGTTTCAAGAAACCCCAATCCGTGGCTGGAGACAGCCAAACCGCGGTGGGTGC 1831

QY 844 CACTGGATGAAATCATGGAGATTGAAGAGGAAAAAGAAAGAACTGCCCGCAACAGACTACT 903
Db 1832 TGGACGAGATCATCAGCAGGAGGAAAGCAAAAGGAGCGGTGCCAACCCGCAAGGACTACT 1891

QY 904 GGCTACAGCCTGAAATTTATTTGTAATTAATTAACCAAGAAACTGGGAGAGAAATATCAT 963
Db 1892 GGCTGCAACAGGATTCGTGTGCTCAAAATTTATTTCCAAATTCATATGGGCGAAAAGTTCTTCA 1951

QY 964 AGAAAAAGGCTATTGTTAAGGAAGTAAATTTGACAAATATACAGCTGTTGTGGAAGATGTTG 1023
```

Db 1952 AACAAAAAGCGTTGCTTGGAGCTAAATGACAGATATCAGGCGCAAAATCAAGTCTTGG 2011
Qy 1024 ATTCTGGAGACAAGCTGAAACTTGACCA 1051
Db 2012 AGACTGGGGAAGGCTAAAAAGTGGATCA 2039
RESULT 7
US-11-097-143-41017
; Sequence 41017, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41017
; LENGTH: 4394
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41017
Query Match 13.7%; Score 209.4; DB 10; Length 4394;
Best Local Similarity 58.6%; Pred. No. 4.3e-40;
Matches 363; Conservative 0; Mismatches 256; Indels 0; Gaps 0;
Qy 64 CCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCACAGCATCAAGTCCAAGG 123
Db 3756 CGATGGTTCGCGCGAGGTAGGTACGCCCAAGTACCTCGCCACAAAGATGAATCGAAG 3815
Qy 124 GGCTGAGAGCTACGCTGATTTGTCAGATGTCAGAGAGAGTCCGCGAGCAGAAATG 183
Db 3816 GTCTGAGAGCTGCGCTGATCTCCAGATGTCGAGAGAGTGCCTGCGGATGAAGACG 3875
Qy 184 GCTTTAAGTGTCAATGATGTCGGAATCTCATCAGAGACAATTAATGCTGGCTTCAGAA 243
Db 3876 GCTTCAAGTCCACACAGTACGAGGAGTCCACAGCGCCAGTGTCTCTCTTTGCGGACA 3935
Qy 244 ATCTCAGAGATTTATGGATTAATTTTCAGAGGAATTCGGAATAGCTTTCTAGAACTTC 303
Db 3936 ATCTGCAAAATCTCTGCACAGCTTCAGCAAGAGTTCCTCCGCGGTACATGGAGTTGC 3995
Qy 304 TCAGGAGACGCTTGGCACTTAAAGGTCACACACATTTGTCTACAGCAATATACATCA 363
Db 3996 TGGCCGCGCGGTTCGGCACCAGAGGAAACCAAGCTACAGATCTACAGAGTACATTC 4055
Qy 364 GCCACCCAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTA 423
Db 4056 CCACAGAGAGCATCCACATGACGCCACCCGATGGCTCACCTGTCTCGACTACGTGA 4115

Qy 424 AGTGGCTGGCAGAGAGGCTTTGTCAAAGTGGACGACACACAAAGGCTGGTATATTC 483
Db 4116 AGTGGCTGGCAGGACTGGCAAGTGATAGCGATGAGACGAGAGGCTGGTTCGTCA 4175
Qy 484 AGTACATAGACAGGAGCCAGAAACTATTCGCGCGCACTGGAACCTGGAGAAAAAGAAA 543
Db 4176 CCTACATTGATCGCAAGTCCAGAGGCCATGGAAACGACAGCGCAAGGCTGATCCCAAGGAGA 4235
Qy 544 AGCAGGACCTTGATGATGAAGAAAAAACTGCCAAAATTTATTGAAGAGCAAGTGAGAAGAG 603
Db 4236 AGATGAGAGAGCAGCAGGAGCGGATGCGCGACTTCATTGAGCAGCAGATTTAAAAATG 4295
Qy 604 GCCTGAAAGGGAAGGAACAGAGAGTCCCTACTTTTACCGGAATTAAGCAGAGAAAAATGATG 663
Db 4296 CCAAGGCCAAGGACGCGGAGGAGGACGAAGGCCAGAGAAGTTTACCGAGCTAAAGCGCG 4355
Qy 664 AAGAGAAAGTCACTTTAA 682
Db 4356 AAGAGAAAGCAACCACTCA 4374
RESULT 8
US-11-097-143-24532
; Sequence 24532, Application US/11097143
; Publication No. US2005020858A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24532
; LENGTH: 4582
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-24532
Query Match 13.7%; Score 209.4; DB 10; Length 4582;
Best Local Similarity 58.6%; Pred. No. 4.4e-40;
Matches 363; Conservative 0; Mismatches 256; Indels 0; Gaps 0;
Qy 64 CCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCACAGCATCAAGTCCAAGG 123
Db 3944 CGATGGTTCGCGCGAGGTAGGTACGCCCAAGTACCTCGCCACAAAGATGAATCGAAG 4003
Qy 124 GGCTGAGAGCTACGCTGATTTGTCAGATGTCAGAGAGAGTGCCTGGAGCAGAAATG 183
Db 4004 GTCTGAGAGCTGCGCTGATCTCCAGATGTCGAGAGAGTGCCTGATGAATCGAAG 4063
Qy 184 GCTTTAAGTGTCAATGATGTCGGAATCTCATCAGAGACAATTAATGCTGGCTTCAGAAA 243
Db 4064 GCTTCAAGTGGCAACAGATGAGCGAGTCCCAAGCGGAGTGTCTCTCTTTGCGGACA 4123

QY 244 ATCCCTCAGCAGTTTATGATTAATTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTC 303
DB 4124 ATCCCTGGCAAAATTCCTGCACACGTTTCAGCAAAAGAGTTCTCCGACGGCTACATGGAGTTGC 4183
QY 304 TCAGGAGAGCGTTTCGGCACTAAAGGGTCCACAAACATTTGTCTACAAAGATACATCA 363
DB 4184 TCGCGCGGGGTTTCGGCAGCAAGCGAAACAGCGCCAAACAAAGATCTACCAAGGAGTACATTG 4243
QY 364 GCCACGAGAGCACATCCACATGAATGCGCACTTCAGTGGGAAATCTCTGACTGATTTTACTTA 423
DB 4244 CCCAAGAGGACACATCCACATGAACGCCACCGGATGGCTCACCCTGTCCGACTACGTGA 4303
QY 424 AGTGCTGGCAGAGAGGCTTTGTGCAAGTGGAGCGACACACAAAGGCTGGTATATTC 483
DB 4304 AGTGCTGGCGGAGTCTGGGCAAGTGATAGCGGATGAGACGAGAGGGGCTGGTTCGTGA 4363
QY 484 AGTACATAGACAGGACCCAGAAACTATCCGCGCGCACTGGAATCTGGAGAAAGAGAAA 543
DB 4364 CCTACATTGATCGCAGTCCAGAGGCCATGGAAACGACAGCGGAAGGCTGATCCAAAGGAGA 4423
QY 544 AGCAGGACCTTGATGATGAAGAAAAAACTGCCCCGCAATTTTATTTGAAGAGCAAGTGAAGAG 603
DB 4424 AGATCGAAGAGCACGACGAGGAGCGATGCGCGACTTCATTGAGCAGCAGATTAAATG 4483
QY 604 GCCTGGAAGGGAAGGAACAGAGAGTCCCTACTTTTACGGAATTAAGCAGAGAAAAATGATG 663
DB 4484 CCAAGGCCAAGACGCGGAGGAGGACGAAGGCCAGGAGAAGTTTACCGAGCTTAAAGCGCG 4543
QY 664 AAGAGAAAGTCACGTTTAA 682
DB 4544 AAGAGAACCAACCACTCAA 4562

RESULT 9

US-10-424-599-25484
; Sequence 25484, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 25484
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MBT3847_123013C.1
US-10-424-599-25484

Query Match 13.5%; Score 205.6; DB 7; Length 777;
Best Local Similarity 62.8%; Pred. No. 1.4e-39;
Matches 319; Conservative 0; Mismatches 189; Indels 0; Gaps 0;
QY 61 TCGCCATGGGGAAGTCGGATTTCTTATCTCCCAAGGCTATCCCAACAGGATCAAGTCA 120
DB 135 TCGTGTAGGGGAAAAATGAGTTTCTCACACCTTAAAGCAATTCGAATCGAATCAAGCAA 194
QY 121 AGGGCTCGAAGACTAGCTGGTATGCGAGATGCGCAGAGAGTCCGGGACGAGA 180
DB 195 AAGGATTGAGAAGCTTCGGTGGTATGCGAGATGTCAGAAAGCAGTCCCGAGATGAGA 254
QY 181 ATGGCTTTAAAGTGTCAATTGTATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTCAG 240
DB 255 ATGGCTTTAAATGCCATTGTCATGAGTGAAGGCCACCAAGCGTCAATGTCAGATTTTGGAC 314

QY 241 AAAATCCTCAGCAGTTTATGATTAATTTTCAGAGGAATTCGAAATGACTTTCTAGAAC 300
DB 315 AAAACCCACACCGGATAGTTGAGGGCTATTTCGGAAGAGTTTGAGAGTACTTTTCTGGAGC 374
QY 301 TTCTCAGGAGAGCGTTTGGCACTAAAGGGTCCACAAACATTTGTCTACAAAGATACATA 360
DB 375 ACATGAAGCGAGTCACCGATTACCGGTGTGGCAGCCACTGTAGTTTATACGNAATACA 434
QY 361 TCAGCCACCGAGAGACATCCACATGAATGCCATCAGTGGGAACTCTGACTGATTTTA 420
DB 435 TAAATGACAGACACCAATTCATTAATGAATCTACTCAGTGGGCTACGCTTACTGAGTTTG 494
QY 421 CTAAGTGGCTGGCAGAGAGGCTTTGCAAAAGTGCAGAGACACCAAAAGGCTGGTATA 480
DB 495 TTAAGTACTTGGGTGCGAACTGGCAAAATGTAAAGTTGAGGAAACACCCAAAGGATGGTTCA 554
QY 481 TTCAAGTACATAGACAGGACCCAGAACTATCCGCGCGCAACTGGAATCTGGAGAAAAAGA 540
DB 555 TTACATATATAGATAGAGATTCAGAAACCCCTTTTCAAGGAGAGGATGAAGAATAAGAAA 614
QY 541 AAAAGCAGGACCTTGATGATGAAGAAAA 568
DB 615 TCAAGGCAGATATGTTAGTAGAAGAAAA 642

RESULT 10

US-10-739-930-2955
; Sequence 2955, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 2955
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-CLUSTER7583_2
US-10-739-930-2955

Query Match 13.4%; Score 204.8; DB 8; Length 1713;
Best Local Similarity 62.9%; Pred. No. 3.4e-39;
Matches 317; Conservative 0; Mismatches 187; Indels 0; Gaps 0;
QY 62 CGCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCCCAACAGGATCAAGTCCAA 121
DB 171 CGCGATGGGGAAGCAGCAGAGTTCTTGACGCGCGAAGCGATCCGAAACCGGCGAA 230
QY 122 GGGCTGCAGAGCTAGCTGGTATTCAGATGTCGAGAGCAGATGCCAGAGCAGTCCCGGACGAGAA 181
DB 231 GGGCTGCAGAGCTGCGGTGGTACTGCCAGATGTGTGAGAGCAGTCCCGGACGAGAA 290
QY 182 TGGCTTTAAGTGTCAATTGTATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTCAGA 241
DB 291 CGGGTTCAAGTGCCACTGCATGTCGAGTCCGACCGGCGAGATGCAAGGTGTTCCGCAT 350
QY 242 AAATCCTCAGCAGTTTATGATTAATTTTCAGAGAAATTCGAAATGACTTTCTAGAACT 301
DB 351 GCGCGCCGACCGCTGCTGCGAGGGCTTCTCCGAGGAAATTCCTCAGATGCTTTCTCTCCCT 410
QY 302 TCTCAGGAGAGCTTTTGGCACTAAAGGGTCCACAAACATTTGTCTACAAAGATACAT 361
DB 411 CATCCGCGCGGCAACCGCACTCCCGCTGCCCGCACCGTTGTCTACAAAGATACAT 470
QY 362 CAGCCACCGGAGAGCACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTAC 421
DB 471 CGCGCAGCGGACCAAGCTCCACATGAACTCTACGCGGTGGGCCCGCTCACCGAGTTGCT 530

QY 422 TAAGTGGCTGGCAGAGAGGCTTGTGCAAACTGGACGAGACCAACAAAGGCTGGTATAT 481
DB 531 CAAGCTCTCTGGGGCGGAAGGGTACTGCAAGGTTGAGGACACGCCAAGGGGTGTTCTAT 590
QY 482 TCAGTACATAGACAGGACCCAGAAACTATCCGCGGCAACTCGGAACCTGGAGAAAGAA 541
DB 591 GACCTACATCGACGCTGACTCGGACAGGCCCTCAAGGACCCGCTCAAGCGCAAGAGAT 650
QY 542 AAAGCAGGACCTTGATGATGAAGA 565
DB 651 CAAGTCAGACATGGCTGAGGACGA 674

RESULT 11

US-10-425-115-157125
; Sequence 157125, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 157125
; LENGTH: 2046
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_74879C.1
US-10-425-115-157125

Query Match 13.3%; Score 203.2; DB 8; Length 2046;
Best Local Similarity 62.7%; Pred. No. 9.3e-39;
Matches 316; Conservative 0; Mismatches 188; Indels 0; Gaps 0;
QY 62 CGCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAA 121
DB 294 CGCATGGGGAAGCAGAGTCTCTGACGCGGAAGGCGATCGCAATCGGATCAAGCGGA 353
QY 122 GGGGCTGCAGAACTACGCTGGTATTGCCAGATGTGCCAAGCAGTGTCCGGGACGAGAA 181
DB 354 GGGGCTGCAGAACTGCGGTGGTACTGCCAGATGTGTCAAGCAGTGTCCGCGCAGAGAA 413
QY 182 TGGCTTTAAGTGTGATGTATGTCGGAATCTCATCAGAGACACTATTGCTGGCTTCA 241
DB 414 CGGGTTCAAGTGCACCTGCAATGCGGAGTGCACACGCGCAGATGCAAGTGTTCGGCAT 473
QY 242 AAATCTCTCAGCAGTTTATGATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACT 301
DB 474 GGGCGCGCAGCGCTGCTCAGGGCTTCTCCGAGGAATCTCTCGAGTCTTCTCTCCCT 533
QY 302 TCTCAGGAGACGCTTTGGCACTAAAAGGGTCCACAAACATTTGTCTAACAGGAATACAT 361
DB 534 CATCCGCGCGCGCACCGCACTCCCGCTGCGCGCACCGTGTGTCTAACAGGATACAT 593
QY 362 CAGCCACCGAGAGACATCCACATGAATGCACTCAGTGGGAACCTCTGACTGATTTTAC 421
DB 594 CGCGGACCGGACACCATCTCCACATGAACCTCTACGCGGTGGGCGCACCGCTCACCGAGTTCGT 653
QY 422 TAAGTGGCTGGGACGAAGGCTTGTGCAAGTGGACGAGACACCAAAAGGCTGGTATAT 481
DB 654 CAACTCTCTGGGGCGCAAGGGTACTGCAAGTTGAGACACGCCCAAGGGGTGTTCTAT 713
QY 482 TCAGTACATAGACAGGACCCAGAACTATCCGCGGCAACTGTGAACTGGAGAAAGAA 541
DB 714 GACCTACATCGACCTGACTCGGAGCAGGCGCTCAAGGACCGCTCAAGCGCAAGAGAT 773
QY 542 AAAGCAGGACCTTGATGATGAAGA 565

RESULT 13
US-10-425-114-22614
; Sequence 22614, Application US/10425114

DB 774 CAAGTCAGACATGGCTGACGACGA 797
RESULT 12
US-10-425-114-2975
; Sequence 2975, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 2975
; LENGTH: 1614
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700238628_FLI
US-10-425-114-2975

Query Match 13.2%; Score 201.6; DB 7; Length 1614;
Best Local Similarity 62.5%; Pred. No. 2e-38;
Matches 315; Conservative 0; Mismatches 189; Indels 0; Gaps 0;
QY 62 CGCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAA 121
DB 76 CACGATGGGGAAGCAGAGTTCCTGACGCGGAAGGCGATCGCAACCGGATCAAGCGGAA 135
QY 122 GGGGCTGCAGAAAGTACGCTGGTATTGCCAGATGTGCCAAGCAGTGTCCGGGACGAGAA 181
DB 136 GGGGCTGCAGAAAGTCCGGTGGTACTGTCAAGTGTGTCAAGCAGTGTCCGCGCAGAGAA 195
QY 182 TGGCTTTAAGTGTGATGTATGTCGGAATCTCATCAGAGACACTATTGCTGGCTTCA 241
DB 196 CGGGTTCAAGTGCACCTGCAATGTCGGAGTGCACACGAGGCGAGATGCAAGTGTTCGGCAT 255
QY 242 AAATCTCTCAGCAGTTTATGATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACT 301
DB 256 GGGCGCGCAGCGCTGCTCAGGGCTTCTCCGAGGAGTTCCTCGAGTCTTCTCTCCCT 315
QY 302 TCTCAGGAGACGCTTTTGGCACTAAAAGGGTCCACAAACATTTGTCTAACAGGAATACAT 361
DB 316 CATCCGCGCGCGCACCGCACTCCCGCTGCGCGCACCGTGTCTAACAGGATACAT 375
QY 362 CAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTAC 421
DB 376 CGCGGACCGCGCACCGCTCCACATGAACCTCCACGCGTGGGCGCACCGCTCACCGAGTTCGT 435
QY 422 TAAGTGGCTGGGACGAAGAGGCTTGTGCAAGTGGACGAGACACCAAAAGGCTGGTATAT 481
DB 436 CAAAGTCTCTGGGCGCGGAGGGTACTGTAAAGTTCGAGGACACGCCCAAGGGGTGGTTCAT 495
QY 482 TCAGTACATAGACAGGACCCAGAAACTATCCCGCGCAACTGGAACTGGAGAAAGAA 541
DB 496 GACCTACATCGACCTGACTCGGAGCAGGCGCTCAAGGACCGCTCAAGGACCGCTCAAGCGCAAGAGAT 555
QY 542 AAAGCAGGACCTTGATGATGAAGA 565
DB 556 CAAGTCAGACATGGCTGACGACGA 579

```

; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 22614
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI
US-10-425-114-22614

Query Match      13.2%; Score 201.6; DB 7; Length 1713;
Best Local Similarity 62.5%; Pred. No. 2.1e-38;
Matches 315; Conservative 0; Mismatches 189; Indels 0; Gaps 0;

Qy      62  CGCCATGGGGAAGTCGGATTTCCTTACTCCCAAGGCTATCGCAACAGAGATCAAGTCCAA 121
Db      235  CACATGGGGAAGCAGAGTTCTCGCCGAGCGGATCGCAACCGGATCAAGCGCA 294

Qy      122  GGGCTGCAGAAGCTACGCTGGTATTTGCCAGATGTCAGAGAGAGTGCCTGGGACGAGAA 181
Db      295  GGGCTGCAGAAGCTGCGGTGGTACTGTCTGATGTCAGAGAGAGTGCCTGGGACGAGAA 354

Qy      182  TGGCTTTAAGTGTCTATTGTATGTCCGAATCTCATCAGAGACAACTATTGCTGGCTTCAGA 241
Db      355  CGGGTTCAAGTGCCTACTGCATGTCCGAGTGCAGCAGAGGAGATGCGAGGTTCGCGCAT 414

Qy      242  AAATCCTCAGCAGTTATATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAATCT 301
Db      415  GGCCCCGACCGTCTGCGAGGGTTCCTCGAGAGATTCCTCGAGTCTCTCTCTCTCTCTCT 474

Qy      302  TCTCAGGAGACGCTTTGGCACTAAAGGTTCCAAACAACTATTTGTCTACACGAAATACAT 361
Db      475  CATCGCCGCGCGCACGCCCACTCCCGCGTGCSCGCCACCGTCGTCTACACGAGTACAT 534

Qy      362  CAGCCACGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTAC 421
Db      535  CGCCGACCGGACCAACGCTCCATGAATCCACGCGGTGGGCGACGCTCACCGAGTTCGT 594

Qy      422  TAAGTGGCTGGCAGAGAAGGCTCTGCAAGTGCACGAGACACCAAAAGGCTGGTATAT 481
Db      595  CAAGTTCCTGGGGCGAGGGGTACTGTAAGTTCAGGACACGCCCAAGGGGTGGTTCAT 654

Qy      482  TCAGTACATAGACAGGGACCCAGAAACTATCCGCGGCAACTGTGGAACCTGGAGAAAGAA 541
Db      655  GACCTACATCGACCGTGACTCGGACAGGCCGTCAAGGACGCCCTCAAGCGCAAGGAT 714

Qy      542  AAAGCAGGACCTTGATGATGAAGA 565
Db      715  CAAGTCAGACATGGCTGACGACGA 738

```

```

; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 89433
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1
US-10-437-963-89433

Query Match      12.6%; Score 192; DB 7; Length 538;
Best Local Similarity 65.8%; P: 0.24e-36;
Matches 279; Conservative 145; Indels 0; Gaps 0;

Qy      54  GCCGTGTCGCCATGGGGAAGTCGAGATTCTTACTCCCAAGGCTATCCCAACAGGATC 113
Db      112  GTCGGCGAAGAGATGGGGAACACAGAGTTCCTGACCGCAGGCGATCCGAACAGGATC 171

Qy      114  AAGTCCAAAGGGGCTGCAGAACTAGCTGTGGTATTCGCCAGATGTCCAGAGCAGTCCCGG 173
Db      172  AAGCGAAGGGGCTGCAGAACTGCGGTGGTACTGCCAGATGTCCAGAAAGCAGTCCCGC 231

Qy      174  GACGAGAAATGGCTTTAAGTGTCAATGTATGTCGGAATCTCATCAGAGACAACTATTGCTG 233
Db      232  GACGAGAAATGGCTTTCAAGTGCACCTGCATGTGCGAGTGCACAGCGCCAGATGCAGGTG 291

Qy      234  GCTTCAGAAATCTTCAGCAGTGTATGGATTATTTTCAGAGGAAATCCGAAATGACTTT 293
Db      292  TTCGGCCAGGCCCCCGACCGAGTGTGTCAGGGCTTCTCCGAGGAGTTCCTCGACGCTTC 351

Qy      294  CTAGAACTTCTCAGGAGACGCTTTGGCACTAAAGGGTCCCAACAACTTGTCTACAAC 353
Db      352  CTCACCTTGTCTCCGCGCGCCACCGACACTCCCGCATCCGCGCACCGTCTGTCTACAAC 411

Qy      354  GAATACATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACT 413
Db      412  GAGTTCATCGCGGACCGCCACACAGTCCACATGAATCCACGCGTGGGCGCACGCTCACC 471

Qy      414  GATTTTACTAAGTGGCTGGGCGAGAGAGGCTTGTGCAAGTCGAGAGACACCAAGAGGC 473
Db      472  GAGTTCAGTTCAGTTCCTCGGCGCGAGGGCCACTGCAAGTTTGAGAGACTCCCAAGGG 531

Qy      474  TGGT 477
Db      532  TGGT 535

RESULT 15
US-10-767-701-25587
; Sequence 25587, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 25587
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:

```

OTHER INFORMATION: Clone ID: 30963818
US-10-767-701-25587

Query Match 10.3%; Score 156.8; DB 7; Length 584;
.Best Local Similarity 64.7%; Pred.No. 1.1e-27;
Matches 233; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy	66	ATGGGGAAGTCGGATTCTTCTACTCCCAAGGCTATGCCCAACAGGATCAAGTCCAAGGGG	125
Db	225	ATGGGGAAGTCGGATTCTTCTACTCCCAAGGCTATGCCCAACAGGATCAAGTCCAAGGGG	284
Qy	126	CTGCAGAACTACGCTGGTATTGCCAGATGTGCCAGAGCAGTGCCTGGGACGAGAAATGGC	185
Db	285	CTGCAGAACTACGCTGGTATTGCCAGATGTGCCAGAGCAGTGCCTGGGACGAGAAATGGC	344
Qy	186	TTTAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAACTATTGTGCTTTCAGAAAT	245
Db	345	TTCAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAACTATTGTGCTTTCAGAAAT	404
Qy	246	CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTC	305
Db	405	CCCGACCGCGTCGTCCGAGGAGTTCCTCGAGTCTCTCTCCCTCATC	464
Qy	306	AGGAGACGCTTTGGCCTAAAGGGTCCACAAACATTTGTCTACAAACGAATACATCAGC	365
Db	465	CGCCGCGCGCACCGCCACTCCCGCGTCGCGCCACCGTCTCTACAAACGAGTACATCGCG	524
Qy	366	CACCGAGAGACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAG	425
Db	525	GACCGTCACCACTCCACATGAATCCACCGCGTGGGCCACGCTCACCGAGTTCGTCAAG	584

Search completed: November 27, 2005, 02:22:24
Job time : 1765.53 secs

This Page Blank (uspic)

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	44	2.9	437	1	US-10-821-234-402	Sequence 402, App
2	40.4	2.6	1575	7	US-11-074-176-145	Sequence 145, App
3	39.4	2.6	3254	1	US-10-793-626-4202	Sequence 4202, App
C	39	2.6	2970	1	US-10-793-626-2593	Sequence 2593, App
	39	2.6	3549	1	US-10-793-626-3796	Sequence 3796, App
C	37.8	2.5	966	1	US-10-793-626-1525	Sequence 1525, App
	37.8	2.5	3742	1	US-10-793-626-4076	Sequence 4076, App
7	37.8	2.5	3782	1	US-10-793-626-3753	Sequence 3753, App
8	37.8	2.5	3985	1	US-10-793-626-3856	Sequence 3856, App
9	37.2	2.4	857	9	US-11-033-764-56	Sequence 56, Appl
11	37.2	2.4	3927	1	US-10-793-626-4223	Sequence 4223, App
12	37.2	2.4	4189	1	US-10-793-626-3898	Sequence 3898, App
13	37.2	2.4	48000	7	US-11-159-597-20	Sequence 20, Appl
14	37	2.4	345	1	US-10-793-626-2251	Sequence 2251, App
15	37	2.4	3153	1	US-10-793-626-3415	Sequence 3415, App
C	37	2.4	3555	1	US-10-793-626-3996	Sequence 3996, App
	37	2.4	3666	1	US-10-793-626-3543	Sequence 3543, App
17	36	2.4	340000	7	US-11-102-978-3	Sequence 3, Appl
18	36	2.4	340000	7	US-11-102-978-3	Sequence 3, Appl
C	35.8	2.3	2913	1	US-10-793-626-3464	Sequence 3464, App
	35.8	2.3	4017	1	US-10-793-626-4104	Sequence 4104, App
C	35.6	2.3	3041	1	US-10-793-626-4368	Sequence 4368, App
	35.6	2.3	3509	7	US-11-077-550-19	Sequence 19, Appl
22	35.6	2.3	340000	7	US-11-102-978-3	Sequence 3, Appl
23	35.6	2.3	340000	7	US-11-102-978-3	Sequence 3, Appl

RESULT 2

US-11-074-176-145
; Sequence 145, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Kluenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5031-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; PRIOR FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 145
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Lactobacillus acidophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1575)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: ORF 1524; Lisk - Putative histidine kinase/may be
; OTHER INFORMATION: involved in stress response
US-11-074-176-145

Query Match 2.6%; Score 40.4; DB 7; Length 1575;
Best Local Similarity 53.9%; Pred. No. 0.57; Indels 0; Gaps 0;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 915 GAAATTTATTTGTAATAACCAAGAACTGGGAGAGAAATATCATAGAAAAAGGCT 974
DB 1003 GAAATGCTTGATTTAACTAGAGCTGAACAAATGATGTTCAATATCCITATGAAGTAACA 1062
QY 975 ATTGTTAGGAAGTAATTCGACAAATATACAGCTGTTGGAAGATGATTTCTGGAGAC 1034
DB 1063 AATGTTAAAGAAACAGTAAACAGAGTTGTTCTGATTTGGCAATGGTTCAATTCAGACTTT 1122
QY 1035 AAGCTGAACCTTGACCAGACTCATTTAGAGACAG 1068
DB 1123 AAGATTCAACTTGATGAAGATGATTTACCAACCAG 1156

RESULT 3

US-10-793-626-4202
; Sequence 4202, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4202
; LENGTH: 3254
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence

US-10-793-626-4202

Query Match 2.6%; Score 39.4; DB 1; Length 3254;
Best Local Similarity 47.7%; Pred. No. 1.5;
Matches 115; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
QY 914 TGAATTTATTTGTAATAACCAAGAACTGGGAGAGAAATATCATAGAAAAAGGC 973
DB 34 TGGAAATAAGGTCGATGTTTCATAGCATGTTCCAGTAGGACAAGATCCACATGAATATGA 93
QY 974 TATTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTTCTGGAGA 1033
DB 94 GGTAAACCTAAAGATATTAAGCATTTAAACAGATGCTGACGTTGTTTATATATGTTT 153
QY 1034 CAAGCTGAACTTGACCAGACTCATTTAGAGACAGTAATTCACGACACAGAAAAAGAA 1093
DB 154 AAACCTAGAACTGGAAATGTTGTTTGAAAAAGCACTTGACCAAGCAGGAAAAATCAAC 213
QY 1094 TCTAGTTTAAATGGAGGCTACAGAGGAATGAAGGTACCTAGATCCATCAATGAGAA 1153
DB 214 AAAAGATAAAAAATGTAGCAGCATCAAAATTAATGTTTAAACCAATATATATCTTAAATGGTGA 273
QY 1154 G 1154
DB 274 G 274

RESULT 4

US-10-793-626-2593/c
; Sequence 2593, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2593
; LENGTH: 2970
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2593

Query Match 2.6%; Score 39; DB 1; Length 2970;
Best Local Similarity 46.2%; Pred. No. 1.8;
Matches 129; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 1249 TTTGAAAAATTTGTTAAACATACATTAATAATCTTAAAGCATCAAAATGTTGTCGCAAGG 1308
DB 2444 TTGTTAAATTCCTCTATTATATATTTTAAACCTTAATTCACCAAGAAATCGTTTCATCTA 2385
QY 1309 CATTATGAGACTCTACTGTGTTAGGTATATTTCTTTGTATATAAAACAAACAGGTTTTTGA 1368
DB 2384 TTATATTATCTTTTAAACCGGAAGTAATTTTATGCAITATAGATATTTGGTACATTA 2325
QY 1369 AAATATTACTGTATAGTTGTTTCAGCTAACTTTGAGAGAAATTAATTTATGTTCTCATGAG 1428
DB 2324 GTTTTAAGTACAATTTGTTTAAAAAACTTACCATAGTTTCATATAATAAATAATTCATTTCTA 2265
QY 1429 GTATCAAACTATGTAATTTTCTCTGTTATTTTGTTCCTTTGTTTCTTACTTATGATGA 1488
DB 2264 TAATTATATCTGAAATAGATTCCTTGATAACTTGTGTACCTTCATCATCTTTTCCAAA 2205
QY 1489 GTTTATATCTTCAATAAAGAAATGTTATTATATAAAAAAAA 1527
DB 2204 GATTGATGATTCCTTCCTGGTTGTTGTTTAAACAAATAAA 2166

```
RESULT 5
US-10-793-626-3796
; Sequence 3796, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3796
; LENGTH: 3549
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3796

Query Match      2.6%; Score 39; DB 1; Length 3549;
Best Local Similarity 46.2%; Pred. No. 2;
Matches 129; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 1249 TTGGAATTTGTTAAACAATACATATAAAATCTTAAAGCATCAAAATGGTGGTCCCAAGG 1308
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1309 CATTATGAGACTCTACTGTGTAGGATATATCTTTTGTATATAAAACAACAGTTTTTGA 1368
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2290 TTATATTATCTTTTAAACGGAAGTAAATTTTATGATATATAGATATGTCATATTA 2349
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1369 AATATATCTGTATAGTTGTCAGTAAACTTTGAGAAGAAATTAATTATGCTCATGAG 1428
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2350 GTTTTAAGTACAAATTTGTTTAAAACTTACCATAGTTTCATATAAAATAATTCATTCTA 2409
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1429 GTATCAAACTATGTAATTTGTCTTGTATATTTTGTCTTGTGTAATTTACTTCATGA 1488
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2410 TAATTATATCTGAATAGATTCCTTGATGAACCTGTTTACCTTCATCATCTTTTCCAA 2469
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1489 GTTTATATCTTCAATTAAGAAATGTTATTATAAAAAA 1527
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2470 GATTGATGATCTTCTCTGTTGTTTAAACAATAAA 2508
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
US-10-793-626-1525/c
; Sequence 1525, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1525
; LENGTH: 966
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1525
```

```
Query Match      2.5%; Score 37.8; DB 1; Length 966;
Best Local Similarity 53.8%; Pred. No. 1.9;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1368 AAAATATTACTGTATAGTTGTTTCAGCTAAACTTTGAGAAAGAAATTTAATTATGCTCATGA 1427
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 776 ATAATTCTACTGTTTGTAGTACACCTAATGATCGATAAACTCTATCCAGTGTCTCTACA 717
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1428 GGTATCAAACTATGTAATTTGTCTTGTATATTTTGTCTTGTGTAATTTACTTTGATG 1487
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 716 GAATGTGTTTATCAAGTCTTCTTATTTGCTTTTCTTCAATTATATTAATTGATTACA 657
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1488 AGTTTATATCTTCATTAAAGAAATGT 1512
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 656 ACTTCAGTTAAGTTATCGATAATGT 632
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
US-10-793-626-4076
; Sequence 4076, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4076
; LENGTH: 3742
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4076

Query Match      2.5%; Score 37.8; DB 1; Length 3742;
Best Local Similarity 53.8%; Pred. No. 4;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1368 AAAATATTACTGTATAGTTGTTTCAGCTAAACTTTGAGAAAGAAATTTAATTATGCTCATGA 1427
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2536 ATAATTCTACTGTTTGTAGTACACCTAATGATCGATAAACTCTATCCAGTGTCTCTACA 2595
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1428 GGTATCAAACTATGTAATTTGTCTTGTATATTTTGTCTTGTGTAATTTACTTTGATG 1487
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2596 GAATGTGTTTATCAAGTCTTCTCTTATTTGCTTTTCTTCAATTATATAATTGATTACA 2655
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1488 AGTTTATATCTTCATTAAAGAAATGT 1512
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2656 ACTTCAGTTAAGTTATCGATAATGT 2680
Db      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
US-10-793-626-3753
; Sequence 3753, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3753
; LENGTH: 3782
```

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3753

Query Match      2.5%; Score 37.8; DB 1; Length 3782;
Best Local Similarity 53.8%; Pred. No. 4.1;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1368 AAAATATTACTGTAGTGTGTGACCTAACTTCGAGAGAAATTTAATTATGCTCTCATGA 1427
Db 1585 ATAATCTTACTGTTTGTAGTACACCTAATGATCGATAAACTCTATCCAGTGTCTCTACA 1644

QY 1428 GGTATCAAACTATGTAATTTGCTCTGTTATTTTGTGTTTCCCTTGTGTAATTTACTTGTATG 1487
Db 1645 GAATGTGTTTATCAAGCTTCTCTATTTGCTTTTCTTCAATTTATATTTGATTACA 1704

QY 1488 AGTTTATATCTTCATTAAGAATGT 1512
Db 1705 ACTTCAGTTAAGTTATCGATAATGT 1729

RESULT 9
US-10-793-626-3856
; Sequence 3856, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3856
; LENGTH: 3985
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3856

Query Match      2.5%; Score 37.8; DB 1; Length 3985;
Best Local Similarity 53.8%; Pred. No. 4.2;
Matches 78; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1368 AAAATATTACTGTATGTTGTTGACCTAACTTCGAGAGAAATTTAATTATGCTCTCATGA 1427
Db 3275 ATAATCTTACTGTTTGTAGTACACCTAATGATCGATAAACTCTATCCAGTGTCTCTACA 3334

QY 1428 GGTATCAAACTATGTAATTTGCTCTGTTATTTTGTGTTTCCCTTGTGTAATTTACTTGTATG 1487
Db 3335 GAATGTGTTTATCAAGCTTCTCTATTTGCTTTTCTTCAATTTATATTTGATTACA 3394

QY 1488 AGTTTATATCTTCATTAAGAATGT 1512
Db 3395 ACTTCAGTTAAGTTATCGATAATGT 3419

RESULT 10
US-11-033-764-56
; Sequence 56, Application US/11033764
; Publication No. US20050244817A1
; GENERAL INFORMATION:
; APPLICANT: Rubin, Donald H.
; APPLICANT: Organ, Edward L.
; APPLICANT: DuBois, Raymond N.
; TITLE OF INVENTION: Mammalian Genes Involved in Viral

; TITLE OF INVENTION: Infection and Tumor Suppression
; FILE REFERENCE: 01123.0004
; CURRENT APPLICATION NUMBER: US/11/033.764
; CURRENT FILING DATE: 2005-01-12
; PRIOR APPLICATION NUMBER: US/10/877,807
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US/09/509,712
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: PCT/US98/21276
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/062,021
; PRIOR FILING DATE: 1997-10-10
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 56
; LENGTH: 857
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1- 857
; OTHER INFORMATION: n = g, a, c or t (u)
US-11-033-764-56

Query Match      2.4%; Score 37.2; DB 9; Length 857;
Best Local Similarity 43.0%; Pred. No. 2.6;
Matches 96; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 435 AGAGAAGCTTTGTGCAAAAGTGGACGACACCAAAAGGCTGTTATTTACGTACATACAC 494
Db 44 ANAGANAGANAGAGNAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 103

QY 495 AGGACCCAGAAACTATCCCGCCCACTGGAACCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 554
Db 104 GAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 163

QY 555 GATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 614
Db 164 NAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 223

QY 615 AAGGAACAGGAGGTCCTCTACTTTTACGGAATTAAGCAGAGAAA 657
Db 224 AANNAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 266

RESULT 11
US-10-793-626-4223
; Sequence 4223, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4223
; LENGTH: 3927
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4223

Query Match      2.4%; Score 37.2; DB 1; Length 3927;
Best Local Similarity 49.4%; Pred. No. 5.9;
Matches 154; Conservative 0; Mismatches 153; Indels 5; Gaps 2;

QY 1220 ATATGAGACATTTCTAAACTTCGCTGAGTTTGAAATTTGTTAAACATACA---TTAAA 1276
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 26, 2005, 12:53:47 ; Search time 5857.21 Seconds
(without alignments)
10694.773 Million cell updates/sec

Title: US-09-555-529-2
Perfect score: 1102
Sequence: 1 atgggcagctcgattttct.....aaaaaaaaaaaaaaaaaaaa 1102

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	GenEmbl:
1: gb_ba.*	
2: gb_in.*	
3: gb_env.*	
4: gb_ov.*	
5: gb_cm.*	
6: gb_pat.*	
7: gb_ph.*	
8: gb_pr.*	
9: gb_ro.*	
10: gb_sts.*	
11: gb_sy.*	
12: gb_un.*	
13: gb_vi.*	
14: gb_htg.*	
15: gb_pl.*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1102	100.0	1102	6 AX003309	Sequence
2	799.2	72.5	1390	6 AX003331	Sequence
3	799.2	72.5	1414	9 MMKIN17	X58472 Mouse KIN17
4	744.2	67.5	1386	9 BC028860	Mus muscu
5	704.6	63.9	1002	6 AX003310	Sequence
6	680	61.7	180882	14 AC114615	Mus muscu
7	680	61.7	182400	14 AC124010	Mus muscu
8	680	61.7	185378	9 AL772367	Mouse DNA
9	648.2	58.8	221547	14 AC119716	AL772367 Mouse DNA
10	611.8	55.5	2710	9 BC058169	Mus muscu
11	535	48.5	144118	14 AC141995	BC058169 Mus muscu
12	510.2	46.3	130262	14 AC166154	AC141995 Rattus no
13	510.2	46.3	140215	9 AC140205	AC166154 Mus muscu
14	475	43.1	1528	6 AX003308	Sequence
15	475	43.1	1528	8 HSAJ5273	AX005273 Homo sapi
16	469	42.6	2069	8 BC017309	Homo sapi
17	333	30.2	333	6 AX003341	Sequence
18	329.2	29.9	1002	6 AX003311	Sequence

19	309	28.0	461	6 CQ684251	Sequence
20	302	27.4	1117	6 CQ722054	Sequence
21	261.6	23.7	333	6 AX003340	Sequence
22	258.8	23.5	1483	5 BC074606	Xenopus t
23	252	22.9	1501	5 BC079988	Xenopus l
24	198.6	18.0	175480	14 AC127074	Rattus no
25	198.6	18.0	247615	14 AC098061	AC127074 Rattus no
26	190	17.2	1372	2 AK113340	AK113340 Ciona int
27	176	16.0	177276	14 AC154247	Mus muscu
28	174	15.8	952	15 AY039912	Arabidops
29	174	15.8	1267	15 AY051011	Arabidops
30	174	15.8	1563	15 AF360132	Arabidops
31	174	15.8	87967	15 AC005223	Arabidops
32	173	15.7	1824	15 AK069396	AK069396 Oryza sat
33	173	15.7	101073	15 AC091302	Oryza sat
34	173	15.7	110000	15 AP008209	Continuation (205
35	173	15.7	152883	15 AC084319	AC084319 Oryza sat
36	172	15.6	760	2 DME6528	AD006528 Drosophill
37	168.8	15.3	1764	6 BD237772	BD237772 Maize KIN
38	167.6	15.2	1780	6 BD237773	BD237773 Maize KIN
39	167.6	15.2	1780	6 BD237774	BD237774 Maize KIN
40	163	14.8	1241	6 CQ613281	Sequence
41	163	14.8	1467	2 AY084184	AY084184 Drosophill
42	163	14.8	3313	6 CQ613280	Sequence
43	163	14.8	4394	6 CQ613259	Sequence
44	163	14.8	4582	6 CQ596774	Sequence
45	163	14.8	31807	14 AC017679	AC017679 Drosophill

ALIGNMENTS

RESULT 1	AX003309	Sequence 2 from Patent WO929845.	1102 bp	DNA	linear	PAT 24-AUG-2000
LOCUS	AX003309	Sequence 2 from Patent WO929845.				
DEFINITION	AX003309	Sequence 2 from Patent WO929845.				
ACCESSION	AX003309	Sequence 2 from Patent WO929845.				
VERSION	AX003309.1	GI:9927126				
KEYWORDS	Mus sp.					
SOURCE	Mus sp.					
ORGANISM	Mus sp.					
REFERENCE	1	Angulo-Mora, J.F. and Mauffrey, P.				
TITLE	Sequences coding for kin17 protein and their applications					
JOURNAL	Patent: WO 929845-A 2 17-JUN-1999;					
	ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)					

FEATURES	Location/Qualifiers
source	1..1102
	/organism="Mus sp."
	/mol_type="unassigned DNA"
	/db_xref="taxon:10095"

ORIGIN	
Query Match	100.0%; Score 1102; DB 6; Length 1102;
Best Local Similarity	100.0%; Pred. No. 5e-200;
Matches 1102; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 ATGGGCAAGTCGAGATTTCTGAGCCCAAGGCCATGCCCAATAGATTAAAGTCCAAAGGG 60
Db	1 ATGGGCAAGTCGAGATTTCTGAGCCCAAGGCCATGCCCAATAGATTAAAGTCCAAAGGG 60
QY	61 CTCAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGTCGCGACGAGATGGC 120
Db	61 CTCAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGTCGCGACGAGATGGC 120
QY	121 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAGCAACTGTGTGCTTCAGAAAC 180
Db	121 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAGCAACTGTGTGCTTCAGAAAC 180

QY 181 CCTCAGCAGTTATGGAATATTTTTCAGAGGAATCCGAAATGACTTTCTGGAACCTCTG 240
DB |||||
181 CCTCAGCAGTTATGGAATATTTTTCAGAGGAATCCGAAATGACTTTCTGGAACCTCTG 240
QY 241 AGCGCAGCGTTGGGCACTAAAAGGTCACAAACATTTGTACAAATGAATACATCAGC 300
DB |||||
241 AGCGCAGCGTTGGGCACTAAAAGGTCACAAACATTTGTACAAATGAATACATCAGC 300
QY 301 CACCGAGAGCACATCCACATGAACGCTCCCAAGTGGGAGACATGACCGACTTTTACCAG 360
DB |||||
301 CACCGAGAGCACATCCACATGAACGCTCCCAAGTGGGAGACATGACCGACTTTTACCAG 360
QY 361 TGGCTGGGAGAGAGGGCTTGTAAAGTACCAAGCTACCCAGTGGGAGACATGACCGAGCGCA 420
DB |||||
361 TGGCTGGGAGAGAGGGCTTGTAAAGTACCAAGCTACCCAGTGGGAGAGCGCA 420
QY 421 GCATCCGGGAAACGGAAAGAGTCTTCACAGAGCTCCGCCAGCTCGCGAAGAGAGAG 480
DB |||||
421 GCATCCGGGAAACGGAAAGAGTCTTCACAGAGCTCCGCCAGCTCGCGAAGAGAGAG 480
QY 481 TCGGCCCTGGATGAGATCATGAGCTCGAAGAGGAAAGAAAGGACCGACCGACAGAC 540
DB |||||
481 TCGGCCCTGGATGAGATCATGAGCTCGAAGAGGAAAGAAAGGACCGACCGACAGAC 540
QY 541 GCCTGTTACAGCGCGGGATCGTTGTGAAATATTAACGAAGAGCTTGGGAGAAATAT 600
DB |||||
541 GCCTGTTACAGCGCGGGATCGTTGTGAAATATTAACGAAGAGCTTGGGAGAAATAT 600
QY 601 CACAAGAAAGGGTCGTTAAGGAAGTGTGACAGGTACACAGCTGTGTTAAGATG 660
DB |||||
601 CACAAGAAAGGGTCGTTAAGGAAGTGTGACAGGTACACAGCTGTGTTAAGATG 660
QY 661 ACTGACTCGGAGACAGGCTGAACTGGACCGAGACTCAATTAGAGACAGTCAATTCGGCC 720
DB |||||
661 ACTGACTCGGAGACAGGCTGAACTGGACCGAGACTCAATTAGAGACAGTCAATTCGGCC 720
QY 721 CCGGGGAAAGGGTCTAGTTTAAATGAGGCTACAGAGGAATGAAGGCACTCTCGAA 780
DB |||||
721 CCGGGGAAAGGGTCTAGTTTAAATGAGGCTACAGAGGAATGAAGGCACTCTCGAA 780
QY 781 TCCATCAATGAGAAGGCTTTTCAGCCAGTAGTCAATTTGAACTGGACCTTTGAAAGGA 840
DB |||||
781 TCCATCAATGAGAAGGCTTTTCAGCCAGTAGTCAATTTGAACTGGACCTTTGAAAGGA 840
QY 841 CCGCAGTTGGAAGGTATCAATATGAAGACATATCTAAACTTGTCTGAGTTTGAATAAT 900
DB |||||
841 CCGCAGTTGGAAGGTATCAATATGAAGACATATCTAAACTTGTCTGAGTTTGAATAAT 900
QY 901 GATAACACACATTTGAACTGTGAAGCATCAAAATGGTGTAGCCAGGCACTGTGTAC 960
DB |||||
901 GATAACACACATTTGAACTGTGAAGCATCAAAATGGTGTAGCCAGGCACTGTGTAC 960
QY 961 TCTACTGTGTAGGGATTTGTTTGTATTAATAAAAAAATCACTATTTAAATACT 1020
DB |||||
961 TCTACTGTGTAGGGATTTGTTTGTATTAATAAAAAAATCACTATTTAAATACT 1020
QY 1021 AGTGAATAGTTGGGTAATTTATAATAAATCTATGTTTTTTTAAAGTAAAAAATAA 1080
DB |||||
1021 AGTGAATAGTTGGGTAATTTATAATAAATCTATGTTTTTTTAAAGTAAAAAATAA 1080
QY 1081 AAAAAAAAAAAAAAAAAAAAAA 1102
DB |||||
1081 AAAAAAAAAAAAAAAAAAAAAA 1102

RESULT 2
AX003331
LOCUS AX003331 1390 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 24 from Patent W0929845.
ACCESSION AX003331
VERSION AX003331.1 GI:9927146
KEYWORDS Mus sp.

ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Angulo-Mora, J. F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 929845-A 24 17-JUN-1999;
ANGULO-MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)
FEATURES
source 1.1390
Location/Qualifiers
/organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"
ORIGIN
Query Match 72.5%; Score 799.2; DB 6; Length 1390;
Best Local Similarity 79.1%; Pred. No. 2.8e-142;
Matches 1099; Conservative 0; Mismatches 3; Indels 288; Gaps 1;
QY 1 ATGGCAAGTCGGATTCTTCAGCCCAAGGCCATCGCAATAGAAATTAAGTCCAAAGG 60
DB |||||
1 ATGGCAAGTCGGATTCTTCAGCCCAAGGCCATCGCAATAGAAATTAAGTCCAAAGG 60
QY 61 CTCAGAAAGCTTCGCTGTACTGCGCAGATGTGCCAAAAGCAATGCCGCGACGAGATGCG 120
DB |||||
61 CTCAGAAAGCTTCGCTGTACTGCGCAGATGTGCCAAAAGCAATGCCGCGCGAGATGCG 120
QY 121 TTTAAGTCTCACTGTATGCTGAAATCTCATCAAGACAACTGTTGCTGCTGCATCAAAAAC 180
DB |||||
121 TTTAAGTCTCACTGTATGCTGAAATCTCATCAAGACAACTGTTGCTGCTGCATCAAAAAC 180
QY 181 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTTCTGGAACCTCTG 240
DB |||||
181 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTTCTGGAACCTCTG 240
QY 241 AGCGCAGCTTTGGCAGCTAAAGGGTCCAAACAACTTGTCTACAATGAATACATCAGC 300
DB |||||
241 AGCGCAGCTTTGGCAGCTAAAGGGTCCAAACAACTTGTCTACAATGAATACATCAGC 300
QY 301 CACCGAGAGCACATCCACATGAACGCTACCCAGTGGGAGACATGACCGACTTTTACCAG 360
DB |||||
301 CACCGAGAGCACATCCACATGAACGCTACCCAGTGGGAGACATGACCGACTTTTACCAG 360
QY 361 TGGCTGGGAGAGGGCTTGTGTAATGATGAGACACCGAAAGGCTGTGTACATTTCAG 420
DB |||||
361 TGGCTGGGAGAGGGCTTGTGTAATGATGAGACACCGAAAGGCTGTGTACATTTCAG 420
QY 388 ----- 387
DB 421 TACATAGACAGAGACCCAGAAAAACATCCGCGCAACTGGAAATTAGAAAAAAGAGAG 480
QY 388 ----- 387
DB 481 CAAGATCTGGACGATGAAGAAAAAACTCCCAAGTTCAATTGAGGACGAGTGAAGAGAGC 540
QY 388 ----- 387
DB 541 CTGGAAGGGAAGAGCAGGAGACACCTGTTTACAGAACTTAGCCGAGAAAAATGAGGA 600
QY 388 ----- 387
DB 601 GAAAAAGTTACGTTCAATCTGAATAAAGAGCGGTGGCTCAGCGGGAGCTACAACTCC 660
QY 388 -----GGTACCAGTGCCTGAAGCTGTGGGAGCGCAGCATCCGGGAAA 432
DB 661 AAGTCAAGCTCTTTGGGACCAAGTGCATGAAGCTGTCTGGGAGCGCAGCATCCGGGAAA 720
QY 433 CGGAAAGAGTCTTCAAGAGCTCCGCGCCAGCTCGGAAAGAAAGAAAGTTCGCGCTCGAT 492
DB |||||
721 CGGAAAGAGTCTTCAAGAGCTCCGCGCCAGCTCGGAAAGAAAGAAAGTTCGCGCTCGAT 780

Db 205 CCTCAGCAGTTATGGATTATTTTTTTCAGAGGAATCCGAAATGACTTTCTGGAACCTTCG 264
QY 241 AGCGGACGCTTTGGCACTAAAAAGGTCCTCAACAACATTTGTCTACAAATGAATACATCAGC 300
Db 265 AGCGGACGCTTTGGCACTAAAAAGGTCCTCAACAACATTTGTCTACAAATGAATACATCAGC 324
QY 301 CACCGAGACCATCCACATGAACGCTACCCAGTGGGAGACACTGACCGCACTTTACCAAG 360
Db 325 CACCGAGACCATCCACATGAACGCTACCCAGTGGGAGACACTGACCGCACTTTACCAAG 384
QY 361 TGGCTGGGAGAGAGGGCTTTGTAA----- 387
Db 385 TGGCTGGGAGAGAGGGCTTTGTAAAGTGGATGAGACACCGAAGGCTGGTACATTCAG 444
QY 388 ----- 387
Db 445 TACATAGACAGAGACCCAGAAACCCTCCGTCGCAACTGGAATTAGAAAAAAGGAAG 504
QY 388 ----- 387
Db 505 CAAGATCTGGACGATGAAGAAAAAACTGCCAAGTTCAATTGAGGAGCAGGTGAGAAGAGGC 564
QY 388 ----- 387
Db 565 CTGGAGGAAAGACGAGGAGACACTGTTTTTACAGAACTTAGCCGAGAAATAGGAA 624
QY 388 ----- 387
Db 625 GAAAAAGTTACGTTCAATCTGAATAAGAGCGGGTGGCTCAGCGGAGCTACAACATCC 684
QY 388 -----GGTACCAGTCGACTGAAGCTGCTGGGAGCGCAGCATCCGGGAAA 432
Db 685 AAGTCAAGCTCTTTGGGACCAAGTGCACTGAAGCTGCTGGGAGCGCAGCATCCGGGAAA 744
QY 433 CGGAAAGAGCTTTACAGAGCTCCGCCAGCTGCGAAGAGAGAGTCCGCCCTTGGAT 492
Db 745 CGGAAAGAGCTTTACAGAGCTCCGCCAGCTGCGAAGAGAGAGTCCGCCCTTGGAT 804
QY 493 GAGATCATGGAGCTCGAAGAGGAAAGAAAAGGACCGCACGACGACGCTGGTTACAG 552
Db 805 GAGATCATGGAGCTCGAAGAGGAAAGAAAAGGACCGCACGACGACGCTGGTTACAG 864
QY 553 CCGGGATCGTTGTGAAATTTAATCAAGAGAGCTTTGGGAGAAATATCACAAAGAA 612
Db 865 CCGGGATCGTTGTGAAATTTAATCAAGAGAGCTTTGGGAGAAATATCACAAAGAA 924
QY 613 GGGGTCGTTAAGGAGTGTACAGAGTACAGAGCTGTGTAAGATGACTCACTCTGA 672
Db 925 GGGGTCGTTAAGGAGTGTACAGAGTACAGAGCTGTGTAAGATGACTCACTCTGA 984
QY 673 GACAGGCTGAAACTGGACAGACTCAATTTAGAGACAGTCAATTCGGGCCCGGGGAAAAGG 732
Db 985 GACAGGCTGAAACTGGACAGACTCAATTTAGAGACAGTCAATTCGGGCCCGGGGAAAAGG 1044
QY 733 GTTCTAGTTTTAAATGGAGGCTACAGAGAAATGAAGGCACCTCTGAAATCCATCAATGAG 792
Db 1045 GTTCTAGTTTTAAATGGAGGCTACAGAGAAATGAAGGCACCTCTGAAATCCATCAATGAG 1104
QY 793 AAGGCTTTTTCAGCCACGATGATCAATTGAAACTGGACCTTCAAGAGGAGCGAGATTGAA 852
Db 1105 AAGGCTTTTTCAGCCACGATGATCAATTGAAACTGGACCTTCAAGAGGAGCGAGATTGAA 1164
QY 853 GGTATTTCAATATGAAGACATATCTAACTTGTCTGAGTTTGAATTTGATAAACAACA 912
Db 1165 GGTATTTCAATATGAAGACATATCTAACTTGTCTGAGTTTGAATTTGATAAACAACA 1224
QY 913 TTGAACCTGTGAAGCATCAAAATGGTGTAGCAAGGCACCTGTGTAACTCTACTGTGTTA 972
Db 1225 TTGAACCTGTGAAGCATCAAAATGGTGTAGCAAGGCACCTGTGTAACTCTACTGTGTTA 1284
QY 973 GGGGATTTCTTTGTATTAATAAAAAAAATCACTATTTAAATACTAGTCAATAGTTG 1032
Db 1285 GGGGATTTCTTTGTATTAATAAAAAAAATCACTATTTAAATACTAGTCAATAGTTG 1344

QY 1033 GGTAAATTTATAATAAATCTATGTTTTTTTAAAGTGAATAAAAAAAAAAAAAAAAAA 1092
Db 1345 GGTAAATTTATAATAAATCTATGTTTTTTTAAAGTGAATAAAAAAAAAAAAAAAAAA 1404
QY 1093 AAAAAAAAAA 1102
Db 1405 AAAAAAAAAA 1414

RESULT 4
BC028860
LOCUS
DEFINITION Mus musculus antigenic determinant of rec-A protein, mRNA (CDNA
clone MGC:18413 IMAGE:3673930), complete cds.
ACCESSION BC028860
VERSION BC028860.1 GI:22135637
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1386)
AUTHORS Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, V.S., Krzywinski, M.I., Skalska, U., Smal, D.E., Scherch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1386)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunnar, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Louleeg, H.,
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LILNL at: <http://image.llnl.gov>
Series: IRAK Plate: 23 Row: h Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: GenomeScan gene

FEATURES		prediction.	
source		Location/Qualifiers	
		1..1386	
		/organism="Mus musculus"	
		/mol_type="mRNA"	
		/strain="FVB/N"	
		/db_xref="taxon:10090"	
		/clone="MGC:18413 IMAGE:3673830"	
		/tissue_type="Mammary tumor; C3(1)-Tag model. Infiltrating ductal carcinoma. 5 month old virgin mouse."	
		/clone_lib="NCI CGAP Mam6"	
		/lab_host="DH10B"	
		/notes="Vector: pCMV-SPORT6"	
gene		1..1386	
		/gene="Kin"	
		/note="synonym: Kin17"	
		/db_xref="GeneID:16588"	
		/db_xref="MGI:96676"	
		20..1195	
		/gene="Kin"	
		/codon_start=1	
		/product="antigenic determinant of rec-A protein"	
		/protein_id="AAH28860.1"	
		/db_xref="GI:22135638"	
		/db_xref="GeneID:16588"	
		/db_xref="MGI:96676"	
		/translation="MGKSDPLSPKAIANRIKSKGLQKRWYCMQCKQCDENGFKCH CMSEHQRLLLASENPQPMDFSEFRNDFLELRRPFGTKRVNNIVYNIYSHR EHTHMNATQWETLDTFKWLGREGKQVETPKGWIQYIDRDPETIRRLQLEKKKK QLDLDEKTAFTVEQVRGLGEQVETPVFELSRENBEKEVTFNLKAGGAGAT TSXSSLSGSAKLGLGASGRKSSQSSAQPAKKKSALDIMELEBEKKRTATD AWLQGVIVKLIITKLGEKXHKKKGVKVIIDRYTAVKWTDSGDLKLDQTHLETVI PAPGKRVLVNGYRGNEGTLSEINEKAFSATIVIEGTPLKGRVREGIQYEDISKLA"	
CDS			
		67.5%; Score 744.2; DB 9; Length 1386;	
		Query Match 77.6%; Pred. No. 8.9e-132;	
		Best Local Similarity 0; Mismatches 13; Indels 294; Gaps 2;	
		Matches 1066; Conservative	
		1 ATGGGCAAGTCGGAATTTCTGAGCCCAAGCCATGCCCAATAGAATTAAGTCAAAAGG 60	
		20 ATGGGCAAGTCGGAATTTCTGAGCCCAAGCCATGCCCAATAGAATTAAGTCAAAAGG 79	
		61 CTCGAGAAGCTTCGCTGGTACTGCGAGATGCGCAAGCAATGCCCGCAGAGATGCG 120	
		80 CTCGAGAAGCTTCGCTGGTACTGCGAGATGCGCAAGCAATGCCCGCAGAGATGCG 139	
		121 TTTAAGTGTCACGTATGCTGAATCTCATCAAGACAACTGTTGCTGGCTTCAGAAAC 180	
		140 TTTAAGTGTCACGTATGCTGAATCTCATCAAGACAACTGTTGCTGGCTTCAGAAAC 199	
		181 CCTCAGCAGTTTATGGAATATTTTTCAGAGAAATCCGAAATGACTTTCTGGAATCTCTG 240	
		200 CCTCAGCAGTTTATGGAATATTTTTCAGAGAAATCCGAAATGACTTTCTGGAATCTCTG 259	
		241 AGGCGACGCTTGGCCTCAAAAGGTCACAAACAACTTGTCTCAATGAATACATCAGC 300	
		260 AGGCGACGCTTGGCCTCAAAAGGTCACAAACAACTTGTCTCAATGAATACATCAGC 319	
		301 CACCGAGACATCCACATGACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAG 360	
		320 CACCGAGACATCCACATGACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAG 379	
		361 TGGCTGGCGCAGAGGGCTTGTGTAAAT-----	387
		380 TGGCTGGCGCAGAGGGCTTGTGTAAATGGATGAGACACCGAAGGCTGTCATTCAG 439	
		388 -----	387
		440 TACATAGACAGACCCAGAAACCATCCGTCGGCAACTGGAAATAGAAAAAGAGAAG 499	
		388 -----	387
		ORIGIN	
		500 CAAGATCTGGACGATGAAGAAAAAACTGCCAAGTTCATTGAGGAGCAGGTGAGAGAGGC 559	
		388 -----	387
		560 CTGGAAGGAAAGAGCAGGAGACACCTGTTTTACAGAACTAAGCCGAGAAATGAGGAA 619	
		388 -----	387
		620 GAAAAAGTTACGTTCAATCTGAATAAAGGAGCGGGTGGCTCAGCGGGAGCTACAACATCC 679	
		388 -----	387
		680 AAGTCAAGCTCTTTGGGACCAAGTGCACCTGAAGCTCTGGGAGCCGAGCATCCGGGAAA 739	
		433 CGGAAAGAGTCTTTCACAGAGCTCCGCCAGCCCTCGCAAGAAAGAAAGTCCGGCCCTGGAT 492	
		740 CGGAAAGAGTCTTTCACAGAGCTCCGCCAGCCCTCGCAAGAAAGAAAGTCCGGCCCTGGAT 799	
		493 GAGATCATCGAGCTCGAAGAGGAAAAAGAACCGCACCGGACAGACGCTCGGTGTACAG 552	
		800 GAGATCATCGAGCTCGAAGAGGAAAAAGAACCGCACCGGACAGACGCTCGGTGTACAG 859	
		553 CCGGGATCGTTGTGAAAAATTATACGAAGAAGCTTTGGGGAGAAATATCACAAAGAGAAA 612	
		860 CCGGGATCGTTGTGAAAAATTATACGAAGAAGCTTTGGGGAGAAATATCACAAAGAGAAA 919	
		613 GGGGTCTGTTAAGGAAGTGAATGACAGGTACACAGCTGTGTTAAGATGACTGACTCTGGA 672	
		920 GGGGTCTGTTAAGGAAGTGAATGACAGGTACACAGCTGTGTTAAGATGACTGACTCTGGA 979	
		673 GACAGCTGAAACTCGACCCAGACTCATTTAGAGACAGTCAATTCGGCCCGGGGAAAGG 732	
		980 GACAGCTGAAACTCGACCCAGACTCATTTAGAGACAGTCAATTCGGCCCGGGGAAAGG 1039	
		733 GTTCTAGTTTTAAATCGAGGCTACAGAGAAATGAAGGCACTCTCGAATCCATCAATGAG 792	
		1040 GTTCTAGTTTTAAATCGAGGCTACAGAGAAATGAAGGCACTCTCGAATCCATCAATGAG 1099	
		793 AGGCTTTTTCAGCCACGATAGTCAATTGAAACTGGACCTTTGAAAGGACGACAGATTGAA 852	
		1100 AGGCTTTTTCAGCCACGATAGTCAATTGAAACTGGACCTTTGAAAGGACGACAGATTGAA 1159	
		853 GGTATTCAATATGAAGACATATCTAAACTTGTCTGAGTTTGTGAAATTTGATAACAACACA 912	
		1160 GGTATTCAATATGAAGACATATCTAAACTTGTCTGAGTTTGTGAAATTTGATAACAACACA 1219	
		913 TTGAAACTGTGAAGCATCAAAATGGTGTAGCCAGGCACTGTGTAACTCTACTGTGTTA 972	
		1220 TTGAAACTGTGAAGCATCAAAATGGTGTAGCCAGGCACTGTGTAACTCTACTGTGTTA 1279	
		973 GGGGATTTGTTTGTATTAAAAAATAAATCATCTATTAAATTAATTAATTAATTAATTAAT 1032	
		1280 GGGGATTTGTTTGTATTAAAAAATAAATCATCTATTAAATTAATTAATTAATTAATTAAT 1333	
		1033 GGTAAATTTATATAAATAAATCTATGTTTTTTTAAAGTGAATAAAAAAATAAAAAA 1085	
		1334 GGTAAATTTATATAAATAAATCTATGTTTTTTTAAAGTGAATAAAAAAATAAAAAA 1386	
		RESULT 5	
		AX003310	
		LOCUS	1002 bp DNA linear PAT 24-AUG-2000
		DEFINITION	Sequence 3 from Patent WO9929845.
		ACCESSION	AX003310
		VERSION	AX003310.1 GI:9927127
		KEYWORDS	Homo sapiens (human)
		SOURCE	Homo sapiens
		ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
		REFERENCE	1
		AUTHORS	Angulo-Mora, J.F. and Mauffrey, P.

TITLE JOURNAL	Sequences coding for kin17 protein and their applications									
	Patent: WO 929845-A 3 17-JUN-1999;									
	ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)									
FEATURES source	Location/Qualifiers									
	1..1002									
	/organism="Homo sapiens" /mol_type="unassigned DNA" /db_xref="taxon:9606"									
ORIGIN	Query Match									
	Best Local Similarity 63.9%; Score 704.6; DB 6; Length 1002;									
	Matches 819; Conservative 0; Mismatches 119; Indels 13; Gaps 3;									
QY	1	ATGGGCAAGTCGGATTTCTGAGCCCAAGGCCATCGCCAATAGAAATTAAGTCCAAAGGG	60							
Db	58	ATGGGGNAGTCGGATTTCTTACTCCCAAGGCTATCGCCACAGGATCAAGTCCAAGGGG	117							
QY	61	CTCCAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCGGACGAGATGGC	120							
Db	118	CTGCAGAGCTACGCTGGTATTGCCAGATGTGCCAAGACGATGCGGGACGAGATGGC	177							
QY	121	TTTAAAGTGTCACTGTATGTCGAATCTCATCAAGACAACTGTGCTGGCTTCAGAAAC	180							
Db	178	TTTAAAGTGTCACTGTATGTCGGAATCTCATCAGACAACTAATGCTGGCTTCAGAAAT	237							
QY	181	CCTCAGCAGTTATGGAATTAATTTTTCAGAGGAATTCGGAATGACTTTTCTGGAATCTTG	240							
Db	238	CCTCAGCAGTTATGGAATTAATTTTTCAGAGGAATTCGGAATGACTTTTCTGGAATCTTG	297							
QY	241	AGCGACGCTTTGGCACTAAAGGGTCCACACATTTGCTACATGAATACATCAGC	300							
Db	298	AGGAGACGCTTTGGCACTAAAGGGTCCACACATTTGCTACAAAGAAATACATCAGC	357							
QY	301	CACCCAGACCATCCACATGAACGCTCCAGTGGGAGACACTGACCGACTTTTACCAG	360							
Db	358	CACCCAGACCATCCACATGAATGCCATCAGTGGGAAACTCTGACTGATTTTACTAAG	417							
QY	361	TGGCTGGGCAGAGAGGGCTTGTGTAAGGTACCACTGACAGTGTGCTGGGGAGCGCA	420							
Db	418	TGGCTGGGCAGAGAGGGCTTGTGCAAA-----AGTGCACTGAGACGATAGGAAGTTCA	471							
QY	421	GCATCGGGAACCGGAAGAGCTTCAACAGAGCTCGGCCAGCCT-----CGGAAGAG	474							
Db	472	GCATCAGTGAACCGAAAGAAATCTTCCACAGAGCTCAACTCAGTCTAAAGAAAGAA	531							
QY	475	AGAAAGTCGGCCCTGGATCAGATCATGGAGCTCGAGAGGAAAGAAAGGACCCAGG	534							
Db	532	AAGAAATCTGCACCTGGATGAATCATGGAGATTGAAGAGGAAAGAAAGAACTGCCGA	591							
QY	535	ACAGAGCCCTGGTTACAGCCGGGGATCGTTGTGAAAATTTATAACGAAGAGCTTGGGAG	594							
Db	592	ACAGACTACTGGCTACAGCCTGAATTAATTTGTGAAATTTATACCAAGAACTGGGAG	651							
QY	595	AAATATCAAGAAAGAAAGGGTTCGTTAAGGAAGTGAATGACAGGTACACAGCTGTGTTA	654							
Db	652	AAATATCAAGAAAGAAAGGGTATTGTTAAGGAAGTAATTTGACAAATATACAGCTGTGTTG	711							
QY	655	AGATGACTGACTCTGGAGACAGGCTGAACTGGACCACTCATTTAGAGACAGTCATT	714							
Db	712	AGATGATTGATTCTGGAGACAGCTGAACTGTGACCACTCATTTAGAGACAGTAAT	771							
QY	715	CCGGCCCCGGGAAAGGGTTCTAGTTTTTAAATGGAGGCTACAGAGGAATGAAGCACT	774							
Db	772	CCAGCACCAAGGAAAGAAATTTCTAGTTTTTAAATGGGGCTACAGAGGAATGAAGTACC	831							
QY	775	CTCGAATCCATCAATGAGAGGCTTTTTCAGCCACGATAGTCATTGAAACTGGAACCTTTG	834							
Db	832	CTAGAATCCATCAATGAGAGAGCTTTTTCAGCTACTATCGTCACTTATGAAACTGGCCCTTA	891							
QY	835	AAAGGACGAGGTGAAGGTATTCAATATGAGACATATCTAAACTTCCTGAGTTTCA	894							

Db	892	AAAGGACGACAGAGTTGAAGGAATTCAAATATGAGACATTTCTAAACTTGCCTGAGTTTCA	951							
QY	895	AAATTTGATACACACACATTCGAAACT-CTGAAGCATCAAAATGCTGTAGC	944							
Db	952	AAATTTGTTAAACAATACCTTTTAAATCTTAAAGCATCAAAATGCTGTTCGC	1002							
RESULT 6 AC114615/c	AC114615 180882 bp DNA linear HTG 09-MAR-2004									
	Mus musculus clone RP24-82119, WORKING DRAFT SEQUENCE, 4 ordered									
	pieces.									
ACCESSION VERSION KEYWORDS SOURCE ORGANISM	AC114615 5 GI:44866278									
	HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.									
	Mus musculus (house mouse)									
REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus. 1 (bases 1 to 180882) Birren,B., Nusbaum,C. and Lander,E. Mus musculus, clone RP24-82119 Unpublished 2 (bases 1 to 180882) Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguski,K., Bouckhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Collinge,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginder,S., Gord,S., Coyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., LaRocque,K., Lamazares,R., Lander,A., Lehotzky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Miengo,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trifilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M. Direct Submission Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA 3 (bases 1 to 180882) Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguski,K., Bouckhalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Miengo,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ranasingh,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M. Direct Submission Submitted (09-MAR-2004) Whitehead Institute/MIT Center for Genome									

TITLE
JOURNAL

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Sep 9, 2002 this sequence version replaced gi:22625488.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em., EMBL; Sw.,
 SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; Information on the WORMPEP
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-119N4 is
 from the RPCI-23 Mouse PAC Library
 constructed by the group of Pieter de Jong.
 For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACe3.6. Qualifiers
 1. 185378
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP23-119N4"
 /clone_lib="RPCI-23"

FEATURES
 source
 Query Match 61.7%; Score 680; DB 9; Length 185378;
 Best Local Similarity 99.3%; Pred. No. 8.7e-120;
 Matches 683; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

ORIGIN
 388 GGTACCAAGTGCATGAAGCTGCTGGGAGCGCAGCATCCGGGAAACGGAAGAGTCTTCA 447
 139979 GGACCAAGTGCATGAAGCTGCTGGGAGCGCAGCATCCGGGAAACGGAAGAGTCTTCA 139920
 448 CAGAGCTCCGCCAGCTCGGAAGAAGAGTGGCCCTCGATGATGATCATGGAGCTC 507
 139919 CAGAGCTCCGCCAGCTCGGAAGAAGAGTGGCCCTCGATGATGATCATGGAGCTC 139860
 508 GAAGAGGAAGAAAGGACCGCAGCAGCAGCAGCCTGTTACAGCCGGGATCGTTGTG 567
 139859 GAAGAGGAAGAAAGGACCGCAGCAGCAGCAGCCTGTTACAGCCGGGATCGTTGTG 139800
 568 AAAATTATTAACGAAGAAGCTTGGGAGAAATATCACAAGAAGAAAGGGTCTGTTAAGGAA 627
 139799 AAATTATTAACGAAGAAGCTTGGGAGAAATATCACAAGAAGAAAGGGTCTGTTAAGGAA 139740
 628 GTGATTGACAGGTACACAGCTGTGGTAAAGATGATGACTCTCGAGACAGGCTGAACCTG 687
 139739 GTGATTGACAGGTACACAGCTGTGGTAAAGATGATGACTCTCGAGACAGGCTGAACCTG 139680
 688 GACCAGACTCATTTAGAGACAGTCAATTCGCGCCCGGGGAAAGGGTCTAGTTTTAAAT 747
 139679 GACCAGACTCATTTAGAGACAGTCAATTCGCGCCCGGGGAAAGGGTCTAGTTTTAAAT 139620
 748 GGAGGCTACAGAGGAATGAAGGACACTCTCGAATCCCATCAATGAGAAGGCTTTTTCAGCC 807
 139619 GGAGGCTACAGAGGAATGAAGGACACTCTCGAATCCCATCAATGAGAAGGCTTTTTCAGCC 139560

QY 808 ACGATAGTCATTGAAACCTGGACCTTTGAAAGGACCCAGAGTTGAAGGTATTCAATATGAA 867
 139559 ACGATAGTCATTGAAACCTGGACCTTTGAAAGGACCCAGAGTTGAAGGTATTCAATATGAA 139500
 QY 868 GACATATCTAAACTTGGCTTGGTTGAAATTTGATACACACATTTGAAACCTGGAAGC 927
 139499 GACATATCTAAACTTGGCTTGGTTGAAATTTGATACACACATTTGAAACCTGGAAGC 139440
 QY 928 ATCAAAATTTGGTGTAGCCAAAGGCACCTGTGTAACCTCTACTGTGTAGGGGATTTGTTTGT 987
 139439 ATCAAAATTTGGTGTAGCCAAAGGCACCTGTGTAACCTCTACTGTGTAGGGGATTTGTTTGT 139380
 QY 988 ATTAAAAAATAAATCATCTATTAAATATCTAGTGAATAGTTGGGTAAATTTATAATA 1047
 139379 ATTAAAAAATAAATCATCTATTAAATATCTAGTGAATAGTTGGGTAAATTTATAATA 139320
 QY 1048 AATCTATGTTTTTTTTTAAGTGTAAAAA 1075
 139319 AATCTATGTTTTTTTTTAAGTGTCTATAA 139292
 Db
 RESULT 9
 AC119716 221547 bp DNA linear HTG 19-NOV-2002
 LOCUS Rattus norvegicus clone CH230-274116, *** SEQUENCING IN PROGRESS
 DEFINITION *** 17 unordered pieces.
 AC119716
 AC119716.6 GI:25092366
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Muridae; Murinae; Rattus.
 1 (bases 1 to 221547)
 Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gabregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C. B., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Loreshuhwa, L., Louissegh, H., Lozada, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martine, E.,
 Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenan, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwaokeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C.,
 Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

Best Local Similarity 89.6%; Pred. No. 4.1e-92;
Matches 614; Conservative 0; Mismatches 55; Indels 16; Gaps 3;
QY 388 GGTACAGTGCACCTGAAGCTGTGGGAGCGCAGCATCCGGGAAACGGAAAGAGTCTTCA 447
Db 121693 GGACCGAGCGCACTGAAGATGCTGGGGAGCACAGCATCCGTGAGCGGAAAGAGTCTTCA 121634
QY 448 CAGAGCTCCGCCAGCTGC---GAGAGAGAGAAGTCCGCCCTGATGAGATCATGGAG 504
Db 121633 CAGAGCTCAGCTCAGCTCGGAGAGAGAGAAGTCTGCCCTGATGAGATCATGGAG 121574
QY 505 CTCGAGAGAGAAAGAAAGAGCGCACGACAGACGCTCGGTTCACGCGGGGATCGTT 564
Db 121573 ATTGAGAGAGAAAGAAAGAGCGCACGACAGATCTTGGTTACAGCTGGGATCTTT 121514
QY 565 GTGAAATATTAAACGAGAGCTGTGGGAGAAATATCACAAGAGAAAGGGGTCTGTTAAG 624
Db 121513 CGGAAATATTAGGAAAGAGCTGTGGGAGAAATACCAAGAGAAAGGGGTCTGTTAAG 121454
QY 625 GAAGTCATTGACAGGTACACAGCTGTGTTAAGATGACTCTCTGGAGACAGGCTGAAA 684
Db 121453 GAAGTCATTGACAGATACACAGCTGTGTTAAGATGACTCTCTGGAGACAGGCTGAAA 121394
QY 685 CTGGACCACTCATTTAGAGACAGTCATTCCGGGCCCGGGGAAAGGTTCTAGTTTAA 744
Db 121393 CTGGACCACTCATTTAGAGACAGTCATTCCAGGCCCGGGGAAAGAGTTCTAGTTTAA 121334
QY 745 AATGGAGGCTACAGAGAAATGAAGCACTCTCGAATCCATCATCAGAGAGGCTTTTCA 804
Db 121333 AACGGAGGCTACAGAGAAATGAAGCACCCCTCGAGTCCATCAATGAAAGACTTTTTCA 121274
QY 805 GCCACATAGTCATTGAAAGCTGACCTTTGAAAGGACGAGAGTTGAAGGTTATCAATAT 864
Db 121273 GCCACATAGTCATTGAAAGCGGACCTTTGAAAGGCGCAGAGTTGAAGGTTATCAATAT 121214
QY 865 GAAGCATATCTAACTTGCTTCAGTTTGAAATTTGATAACACACATTTGAAAGCTGTCA 924
Db 121213 GAAGCATATCTAACTTGCTTCAGTTTGACAAATTTGTAATACACATTTGAAATGTAA 121154
QY 925 AGCATCAAAATGGTGTAGCCAGGCACTGTGTAACTCTACTGTGTAGGGATTTGTTT 984
Db 121153 AGCATCAAAATGGTGTAGCCAGGCACTGTGTAACTCTACTGTGTAGGGATTTGTTT 121094
QY 985 TGTATTAAAAAATAAATCATCTATTAAATACCTAGTGAATAG-----TTGGTAAATTT 1040
Db 121093 -----ATAAAGAAAGTCATCTATTAAATACCACCTGGATAGCTGTTGGTAAATTT 121043
QY 1041 TATAATAAATCTATGTTTTTTTAA 1065
Db 121042 TATAATAAATCTATTTTTTTTAA 121018

RESULT 12
AC166154
LOCUS
DEFINITION Mus musculus chromosome 8 clone RP24-344F2 map 8, WORKING DRAFT
SEQUENCE, 6 unordered pieces.
AC166154
VERSION AC166154.1 GI:71143327
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 130262)
Birren,B., Nuebaum,C. and Lander,E.
Mus musculus chromosome 8, clone RP24-344F2
Unpublished
2 (bases 1 to 130262)
Birren,B., Nuebaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V.,
Bloom,T., Boguslavskiy,L., Boukhalter,B., Camarata,J., Chang,J.,

Chospel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B.,
DeIrellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L.,
Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D.,
Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T.,
Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R.,
Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C.,
Retta,R., Riese,C., Rogov,P., Roman,J., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R.,
Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (26-JUL-2005) Broad Institute of MIT and Harvard, 320
Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Broad Institute of MIT and Harvard
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
----- Project Information
Center project name: L33627
Center clone name: 344_F2
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 128346 bases at least Q40
Consensus quality: 128716 bases at least Q30
Consensus quality: 129103 bases at least Q20
Insert size: 200000; agarose-fp
Insert size: 129762; sum-of-contigs
Quality coverage: 10.2 in Q20 bases; agarose-fp
Quality coverage: 15.7 in Q20 bases; sum-of-contigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 39772: contig of 39772 bp in length
* 39773 39872: gap of unknown length
* 39873 41564: contig of 1692 bp in length
* 41565 41664: gap of unknown length
* 41665 47820: contig of 6156 bp in length
* 47821 47920: gap of unknown length
* 47921 80877: contig of 32957 bp in length
* 80878 106562: gap of unknown length
* 106563 106662: contig of 25585 bp in length
* 106663 130262: contig of 23600 bp in length.
Location/Qualifiers
1. .130262
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="8"
/map="8"
/clone="RP24-344F2"
/clone.lib="RPC1-24 Male Mouse BAC"
misc_feature 1. .39772


```
/note="assembly_fragment
clone_end:SP6
vector_side:left"
gap 39773..39872
misc_feature /estimated_length=unknown
39873..41564
gap /note="assembly_fragment"
41565..41664
misc_feature /estimated_length=unknown
41665..47820
gap /note="assembly_fragment"
47821..47920
misc_feature /estimated_length=unknown
47921..80877
gap /note="assembly_fragment"
80878..80977
misc_feature /estimated_length=unknown
80978..106562
gap /note="assembly_fragment"
106563..106662
misc_feature /estimated_length=unknown
106663..130262
gap /note="assembly_fragment"
clone_end:T7
vector_side:right"

ORIGIN
Query Match 46.3%; Score 510.2; DB 14; Length 130262;
Best Local Similarity 85.2%; Pred. No. 2.3e-87;
Matches 609; Conservative 0; Mismatches 93; Indels 13; Gaps 3;

Qy 388 GGTACAGTGCTGCTGAAGCTGCTGGGAGCGCAGCATCCGGAAACCGGAAGGCTCTCA 447
Db 100396 GGACCAAGTGCTGCTGCTGAGGAGCAGCATCCGGAAATCGGAAGGCTCTCA 100455

Qy 448 CAGAGCTCCGCCAGCTGC---GAAGAAGAAGTCCGCCCTGGATGAGATCATGGAG 504
Db 100456 CAGAGCTCTCCAGCTGCTGAAAGAAGAGAGATGCTGCCCTGGATGAATATAGAG 100515

Qy 505 CTCGAAGAGGAAAGAAAGACCGCACAGACGCTGGTTACAGCCGGGATCGTT 564
Db 100516 ATTAAGAAGAAAGAAAGGACACACAGGAGACTCTGGTTACAGCCGGGAATCAT 100575

Qy 565 GTGAAGATTATACGAAGAAGCTTGGGAGAAATATACAGAAGAAAGGGTCTTAAG 624
Db 100576 GTGAAGCGTATAGGAAGAAGCTTGGGAGAAATATCAAGAAAGAAAGGGTCAATACG 100635

Qy 625 GAAGTGATTGACAGGTACACAGCTGTGTAAGATGACTGACTCTGGAGACAGGCTGAAA 684
Db 100636 GAAGAGATTGACTGATACACAGCTGTGTAAGATGACTGACTCTGGAGACAGGCTGAAA 100695

Qy 685 CTGGACAGAGCTCATTTAGAGACAGTCAATCCGCCCGGGGAAAGGGTTCTAGTTTA 744
Db 100696 CTGGACAGAGCTCATTTAGAGACAGTCAATCCGCCCGGGGAAAGGGTTCTAGTTTA 100755

Qy 745 AATGAGGCTACAGAGAAATGAAGGCTCTCAATCCATCATGAGAGGCTTTTCA 804
Db 100756 AATGAGGCTACAGAGAAATGAAGGCAACCTCAATATACATCAATGAGAGGCTTTTTCG 100815

Qy 805 GCCAGTAGTCAATTGAAGCTGGACCTTTGAAGAGCAGCAGTGTGAAGGTATTCAATAT 864
Db 100816 GCTAAGATAGTCAATTGAAGCTGGACCTTTGAAGAGCAGCAGTGTGAAGGTATTCAATAT 100875

Qy 865 GAAGACATATCTAACTTGTGAGTTTGAAGAAATTTGATAACAACACATTCGAAGTGA 924
Db 100876 AAAGACATATCTAATCTTGTGAGTTTGAAGAAATTTGTTAGAAACACATTCGAATTTGAA 100935

Qy 925 AGCATCAATTTGGTTAGCCAGGCACTGTGAATCTACTGTGTTAGGGGATTTGTTT 984
Db 100936 AGCATCAATTTGATGCTAGTCAAGGCACTTTGTAATCTACTGTGTTAGGGGATTTGTTT 100995

Qy 985 TGTATTAAGAAAGAAAGAAATCATCTATTATTAATCTAGTGAATAGTTGGGTAAATTTATA 1044
Db 100996 TGTATTAAGAAAGAAAGAAATCATCTATTATTAATCTAGTGAATAGTTGGGTAAATTTATA 10444

Db 100996 TGTAT-----ATAAAAAATCATCTATTTAAGTAATACTAAATAATTCGTTAA-----ATA 101045
Qy 1045 ATAAATCTATGTTTTTTTTTAAGTGTAATAAAAAAAGAAAAAAGAAAAAAGAAAAA 1099
Db 101046 AAATCTTATGTTTTTTTTTAAGTGTAATAAAAAAAGAAAAAATTAACAGATCTCTACTACAA 101100

RESULT 13
AC140205 140215 bp DNA linear ROD 29-MAY-2004
LOCUS Mus musculus BAC clone RP24-344F2 from chromosome 8, complete
DEFINITION sequence.
AC140205
VERSION AC140205.2 GI:47131361
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 140215)
Tomlinson, C., Bielicki, L., Cordum, H. and Haakenson, W.
The sequence of Mus musculus BAC clone RP24-344F2
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 140215)
McPherson, J. D. and Waterston, R. H.
Direct Submission
Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 140215)
Wilson, R. K.
Direct Submission
Submitted (12-MAY-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 140215)
Wilson, R. K.
Direct Submission
Submitted (29-MAY-2004) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On May 12, 2004 this sequence version replaced gi:28475441.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
-----
Summary Statistics
Center project name: M_BB0344F02
-----
NOTICE:
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.
MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu
SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org
NEIGHBORING SEQUENCE INFORMATION:
```

This sequence is the entire insert of the clone.

```
FEATURES
    source          Location/Qualifiers
1..140215
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10090"
    /chromosome="8"
    /map="8"
    /clone="RP24-344F2"
    /clone_lib="RPCI-24"
32..81
    /rpt_family="ERVK"
repeat_region
333..426
    /rpt_family="ERVK"
repeat_region
432..600
    /rpt_family="ERVK"
repeat_region
605..763
    /rpt_family="ERVK"
repeat_region
830..1157
    /rpt_family="ERVK"
repeat_region
1158..11387
    /rpt_family="ERVK"
repeat_region
1426..1641
    /rpt_family="ERVK"
repeat_region
1657..11779
    /rpt_family="ERVK"
repeat_region
2071..21147
    /rpt_family="ERVK"
repeat_region
2242..2391
    /rpt_family="ERVK"
repeat_region
2832..2881
    /rpt_family="ERVK"
repeat_region
3133..3226
    /rpt_family="ERVK"
repeat_region
3232..3400
    /rpt_family="ERVK"
repeat_region
3405..33563
    /rpt_family="ERVK"
repeat_region
3630..3957
    /rpt_family="ERVK"
repeat_region
3958..4187
    /rpt_family="ERVK"
repeat_region
4226..4441
    /rpt_family="ERVK"
repeat_region
4457..4579
    /rpt_family="ERVK"
repeat_region
4871..4947
    /rpt_family="ERVK"
repeat_region
5042..5191
    /rpt_family="ERVK"
repeat_region
5632..5681
    /rpt_family="ERVK"
repeat_region
5933..6026
    /rpt_family="ERVK"
repeat_region
6032..6200
    /rpt_family="ERVK"
repeat_region
6205..6363
    /rpt_family="ERVK"
repeat_region
6430..6757
    /rpt_family="ERVK"
repeat_region
8931..9312
    /rpt_family="L1"
repeat_region
9434..9660
    /rpt_family="L1"
repeat_region
9692..9931
    /rpt_family="L1"
repeat_region
9959..10098
    /rpt_family="L1"
repeat_region
10770..10883
    /rpt_family="L1"
repeat_region
11090..11220
    /rpt_family="Alu"
repeat_region
11240..11299
```

```
repeat_region      /rpt_family="L1"
11704..11890
    /rpt_family="ERV1"
repeat_region
11977..12288
    /rpt_family="ERV1"
repeat_region
13460..13519
    /rpt_family="ERV1"
repeat_region
13586..13801
    /rpt_family="RMER10B"
repeat_region
18269..18413
    /rpt_family="RMER10B"
repeat_region
19731..21946
    /rpt_family="ERV1"
repeat_region
21953..23462
    /rpt_family="ERV1"
repeat_region
23875..24058
    /rpt_family="ERV1"
repeat_region
24558..24711
    /rpt_family="ERVK"
repeat_region
25065..25121
    /rpt_family="L1"
repeat_region
26157..26279
    /rpt_family="RMER10A"
repeat_region
27231..27425
    /rpt_family="RMER10A"
repeat_region
27505..27718
    /rpt_family="RMER10A"
repeat_region
27956..28084
    /rpt_family="MaLR"
repeat_region
28245..28470
    /rpt_family="MaLR"
repeat_region
30676..30798
    /rpt_family="MaLR"
repeat_region
31586..31640
    /rpt_family="L1"
repeat_region
31634..32651
    /rpt_family="L1"
repeat_region
33734..33932
    /rpt_family="MER1_type"
repeat_region
35881..36272
    /rpt_family="MaLR"
repeat_region
38599..39351
    /rpt_family="L1"
repeat_region
40422..40554
    /rpt_family="L1"
repeat_region
41164..41266
    /rpt_family="B2"
repeat_region
42685..43336
    /rpt_family="L1"
repeat_region
45098..45157
    /rpt_family="ERV1"
repeat_region
45351..45480
    /rpt_family="ERV1"
repeat_region
46796..47083
    /rpt_family="L1"
repeat_region
47434..47573
    /rpt_family="B2"
repeat_region
47713..48088
    /rpt_family="L1"
repeat_region
48495..48643
    /rpt_family="Alu"
repeat_region
49036..49254
    /rpt_family="MER1_type"
```

```
Query Match          46.3%; Score 510.2; DB 9; Length 140215;
Best Local Similarity 85.2%; Pred. No. 2.2e-87;
Matches 609; Conservative 0; Mismatches 93; Indels 13; Gaps 3;

Qy 398 GGTACCACTGCACCTGAAGCTGCTGGGAGCGCAGCATCCGGGAAACGGAAAGAGTCTTCA 447
    |||
Db 82385 GGACCAAGTGCACCTGAAGCTGCTGGAGGACACAGCATCCGGGAAATGGAAAGGTCTTCA 82444

Qy 448 CAGAGCTCGGCCACCTGTC---GNAGAAGAAGAGTGGCCCTGGATCAGATCATGGAG 504
```


||||| 1146 AATGAGAAGACTTTTTCAGCTACTATCGTCATTGAAACTGGCCCTTTAAAGAGCGCAGA 1205
QY 847 GTTGAAGGTATTCAATATGAAGACATATCTAAACTTGTCTTGAGTTTGAAGTTTGAATAAC 906
Db 1206 GTTGAAGGAATTCATATGAAGACATTTCTAACTTGCCTGAGTTTGAAGTTTGAATAAC 1265
QY 907 AACACATTGAAGACATCAAAATGGTGTGTTAGCCCAAGCACTGTGTAACTCTACT 966
Db 1266 AATACATTAAATCTTAAAGCATCAAAATGGTGTGTTAGCCCAAGCACTTATGAGACTCTACT 1325
QY 967 GTGTTAGGGGATTTCTTGTATTAAAGAAAAAATCAATCTATTAAATACTAGTGA 1026
Db 1326 GTGTTAGGGATATTCTTTTGTATAAACAAACAGGTTTTTGAATAATATTACTGTATAGT 1385
QY 1027 TAGTTGGGTAAATTTATATAAAATCTATGTTTTTTTTTAAAGTGTAAAAAA 1077
Db 1386 TGTTCAGCTAACTTTGAGAGAAATTAATATTATGCTCATGAGGTATCAAA 1436
RESULT 15
HSAJ5273
LOCUS HSAJ5273 1528 bp mRNA linear PRI 13-DEC-2000
DEFINITION Homo sapiens mRNA for Kin17 protein.
ACCESSION AJ005273
VERSION AJ005273.1 GI:3850703
KEYWORDS Kin17 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Kannouche P., Mauffrey P., Pinon-Lataillade G., Mattei M.G.,
Saxain A., Daya-Grosjean L. and Angulo J.F.
TITLE Molecular cloning and characterization of the human Kin17 cDNA
encoding a component of the UVC response that is conserved among
metazoans
JOURNAL Carcinogenesis 21 (9), 1701-1710 (2000)
PUBMED 10964102
REFERENCE 2
AUTHORS Mauffrey P.
TITLE Direct Submission
JOURNAL Submitted (08-Apr-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 -
Fontenay aux Roses Cedex, 92265, FRANCE
REMARK Revised by [3]
AUTHORS 3 (bases 1 to 1528)
Mauffrey P.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 -
Fontenay aux Roses Cedex, 92265, FRANCE
COMMENT On Nov 7, 1998 this sequence version replaced gi:3046739.
Kannouche P., Mauffrey P., Pinon-Lataillade G., Biard D.S.F.,
Angulo J.F.
Sequences codant pour la Proteine kin17 et leurs Applications -
French Patent Nr 97 15536 1:1-50(1997).
FEATURES
source
1..1528
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="male"
/tissue_type="testis"
1..1528
/gene="Kin17"
66..1247
/gene="Kin17"
/codon_start=1
/protein_id="CAA06462.1"
/db_xref="GI:3850704"
/db_xref="GOA:O60870"
/db_xref="UniProt/TREMBL:O60870"
/translation="MGKSDFLTPKAIANRIKSKGLQKLWRYCQMCKQCRDENGPKCH"

ORIGIN
Query Match 43.1%; Score 475; DB 8; Length 1528;
Best Local Similarity 65.4%; Pred. No. 2e-80;
Matches 897; Conservative 0; Mismatches 180; Indels 294; Gaps 2;
QY 1 ATGGGCAAGTCGGATTTCTGAGCCCAAGGCGCATCGCAATAGAAATTAAGTCCAAAGG 60
Db 66 ATGGGAAAGTCGGATTTCTTACTCCCAAGGCTATCGCCCAACAGGATCAAGTCCAAAGG 125
QY 61 CTCAGAAAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAGAAATGGC 120
Db 126 CTGCAGAAAGCTACGCTGGTATTGCCAGATGTGCCAGAGCAGTGCCTGGGACGAGAAATGGC 185
QY 121 TTAAAGTCTCACTGTATCTGAAATCTCATCAAGACAACTGTTCCTGGCTTCAGAAAC 180
Db 186 TTAAAGTCTGATTTGATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTCAGAAAT 245
QY 181 CCTCAGCAGTTTATGGATTATTTTCAGAGGAATTCGAAATGACTTTTCTGAACTTCTG 240
Db 246 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTTCCGAAATGACTTTTCTGAACTTCTC 305
QY 241 AGGCGACGCTTTGGCAGCTAAAGGGTCCACAAACAACTTGTCTACAATGAATACATCAGC 300
Db 306 AGGAGACGCTTTGGCAGCTAAAGGGTCCACAAACAACTTGTCTACAACGAATACATCAGC 365
QY 301 CACCGAGAGCAGCATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAAG 360
Db 366 CACCGAGAGCAGCATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAG 425
QY 361 TGGCTGGGACAGAGGGGCTTGTGTAA----- 387
Db 426 TGGCTGGGACAGAGAGGGCTTGTGCAAAAGTGGACGAGACACCAAAAGGCTGGTATATTACAG 485
QY 388----- 387
Db 486 TACATAGACAGGAGCCAGAAAATCTATCCCGGCAACTGGAACCTGGAGAAAGAAAAG 545
QY 388----- 387
Db 546 CAGGACCTTGATGATGAAGAAAAAACTGCCAAATTTATTGAAGAGCAAGTGAGAGAGGC 605
QY 388----- 387
Db 606 CTGGAAGGAGGAGAAACAGGAGGTCCCTACTTTTACGGAATTAAGCAGAGAAAATGATGAA 665
QY 388----- 387
Db 666 GAGAAAGTCAGCTTTAAATTTGAGTHAAGAGCATGTAGCTATCCGGAGCAACATCTTCC 725
QY 388-----GGTACCAGTGCTGACTGAAGCTGTGGGAGCGCGACATCCGGGAAA 432
Db 726 AAGTCAAGTACTCTCGGACCGAGTGCTGAAAGCAGTAGGAAGTTTACGATCAGTGA 785
QY 433 CGGAAAGAGTCTTACAGAGCTCCGCCAGCCTGC-----GAAGAAGAGAAAGTGGCC 486
Db 786 CGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCA 845
QY 487 CTGGATGAGATCATCGAGCTCGAAGAGGAAAAAGAACCGCACCGGACAGACGCTCG 546
Db 846 CTGGATGAATCATGAGATTTGAAGAGGAAAAAGAAAGAACTGCCGGAACAGACTACTGG 905
QY 547 TTACAGCCGGGATCGTTGTGAAATTTATAACGAAGAGCTTTGGGAGAAAAATATCACAAG 606
Db 906 CTACAGCCTGAAATTTATTGTGAAATTTATACCAAGAAACTGGGAGAGAAATATCATAG 965

Qy	607	AAGAAAGGGTCTTAAGGAAGTATTGACAGCTACACAGCTCTGGTAAAGATGACTGAC	666
Db	966	AAAAAGCTATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGAT	1025
Qy	667	TCTGGAGACAGGCTGAAACTGGACCAAGCTCATTTAGAGACAGTCAATTCGGCCCCGGGG	726
Db	1026	TCTGGAGACAAGCTGAAACTTGACAGACTCAATTTAGAGACAGTAATTCAGCACCAGGA	1085
Qy	727	AAAAGGTTCTGTTTAAATGAGAGCTACAGAGGAATGAAGGCACTCTCGAATCCATC	786
Db	1086	AAAAGAAATTCAGTTTTTAAATGGAGGCTACAGAGGAATGAAGGTACCCTAGAATCCATC	1145
Qy	787	AATGAGAAGGCTTTTTCAGGCACGATAGTCAATTTGAAACTGGACCTTTGAAAGGACGCAGA	846
Db	1146	AATGAGAAGACTTTTTCAGCTACTATCGTCAATTTGAAACTGGCCCTTTAAAGGACGCAGA	1205
Qy	847	GTTGAAGGTATTCAAATATGAAGACATATCTAAACTTGTGAGTTGAAATTTGATAAC	906
Db	1206	GTTGAAGGAATTCAAATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAAAATTTGTTAAC	1265
Qy	907	AACACATTGAAACTGTGAAGCATCAAAATGGTGTAGCCCAAGGCACTGTGTAACCTACT	966
Db	1266	AATACATTAAAACTTTAAAGCATCAAAATGGTGTAGCCCAAGGCAATATGAGACTCTACT	1325
Qy	967	GTGTTAGGGATTGTTTCTGTTATTAATAAAAAAATCATCTATTTAAATACTAGTGAA	1026
Db	1326	GTGTTAGGGTATATCTTTTGTATAAAAACAACAGGTTTTTTGAAATATTTACTGTATAGT	1385
Qy	1027	TAGTTGGGTAAATTTATAATAAAATCTATGTTTTTTTTTAAAGTGTAAAAAAA	1077
Db	1386	TGTTACAGTAAACTTTGAGAAGAAATTTAATTAATGCTCTCATGAGGTATCAAA	1436

Search completed: November 26, 2005, 19:10:49
Job time : 5868.21 secs

This Page Blank (uspic)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 706.843 Seconds
(without alignments)
10390.557 Million cell updates/sec

Title: US-09-555-529-2
Perfect score: 1102
Sequence: 1 atggcgacgtcggtttctt.....aaaaaaaaaaaaaaaaaaaaa 1102

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1102	100.0	1102	2	AX85550
2	799.2	72.5	1390	2	AX85570
3	790	71.7	1458	2	AAQ7936 Murine Ki
4	704.6	63.9	1002	2	AX85551
5	422.2	38.3	1296	2	AX85549
6	385.4	35.0	679	13	ADQ56943 Novel can
7	329.2	29.9	1002	2	AX85552
8	269.2	24.4	591	4	AAH34095
9	190	17.2	1372	12	ADP22451
10	174	15.8	1459	3	AAC47248
11	168.8	15.3	1614	13	ADO84255
12	168.8	15.3	1713	13	ADX47874
13	168.8	15.3	1764	3	AAAS2589
14	167.6	15.2	1713	13	ADT17629
15	167.6	15.2	1780	3	AAAS2591
16	167.6	15.2	1780	3	AAAS2590
17	163	14.8	1241	4	ABL29199
18	163	14.8	3313	4	ABL29198
19	163	14.8	4394	4	ABL29184

	20	163	14.8	4582	4	ABL18194	Ab118194 Drosophil
	21	136	12.3	1032	3	AAC48794	Aac48794 Arabidops
	22	130.8	11.9	549	3	AAC94929	Aac94929 Cat flea
c	23	116.8	10.6	398	13	ACF87901	AcF87901 Human SIR
	24	95.4	8.7	300	3	AAA01559	Aaa01559 Human col
	25	93.6	8.5	270	6	AAJ73715	Ab173715 Corn tass
	26	85	7.7	696	3	AAC38477	Aac38477 Arabidops
	27	67.4	6.1	431	3	AAC48801	Aac48801 Arabidops
c	28	66.2	6.0	375	5	ABV4911	Abv4911 Human pro
	29	65	5.9	65	6	ABN51894	Abn51894 Mouse spl
c	30	63	5.7	396	13	ACN63020	Acn63020 Cotton ca
	31	62.6	5.7	552	13	ACN45312	Acn45312 Cotton pr
c	32	61.2	5.6	723	5	ADI171984	Adi171984 Human ova
	33	61.2	5.6	723	5	ADL37134	Adl37134 Human ova
c	34	59.4	5.4	390	5	ADL43918	Adl43918 Human ova
	35	59.2	5.4	327	5	ABV08167	Abv08167 Human pro
c	36	59	5.4	2229	2	AAQ23462	Aaq23462 Human B-r
	37	59	5.4	2229	10	AAQ64082	Aaq64082 DNA #3 re
	38	59	5.4	2229	12	ADL16255	Adl16255 Human B-r
	39	59	5.4	2510	2	AAQ21931	Aax21931 Human B-r
	40	59	5.4	2510	6	ABK84128	Abk84128 Human CDN
	41	59	5.4	2510	6	ABL57049	Ab157049 Human pro
	42	59	5.4	2510	6	AAU44844	Aad44844 Human B-r
	43	59	5.4	2510	9	ADA02937	Ada02937 Human BRA
	44	59	5.4	2510	10	ADB72675	Adb72675 Human BRA
	45	59	5.4	2510	10	ADC85417	Adc85417 Human Bra

ALIGNMENTS

RESULT 1

AX85550

ID AX85550 standard; cDNA; 1102 BP.

XX AC AX85550;

XX DT 07-SBP-1999 (first entry)

XX DE cDNA of a gene coding for a mouse deleted kin17 protein.

XX KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
XX KW protein interaction; curved DNA; HIV replication; HIV integration;
XX KW repair enzyme; ss.

XX OS Mus sp.

XX PN FR2772046-A1.

XX PD 11-JUN-1999.

XX PF 09-DEC-1997; 97FR-00015536.

XX PR 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX DR WPI; 1999-359999/31.

XX PT New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.

XX PS Claim 4; Page 31; 69pp; French.

XX CC The present sequence encodes a mouse kin17 protein with amino acids 129-228 deleted. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment

CC can be used to block replication of HIV or its integration into the human
CC genome or to target repair enzymes to curved DNA sites. Expression
CC vectors for kin17 can be used for controlling cell proliferation
XX
SQ Sequence 1102 BP; 373 A; 205 C; 272 G; 252 T; 0 U; 0 Other;

Query Match 100.0%; Score 1102; DB 2; Length 1102;
Best Local Similarity 100.0%; Pred. No. 1.5e-208;
Matches 1102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCCAGTCCGATTTCTGAGCCCAAGGCATCGCAATAGATTAAGTCCAAAGG 60
DB |||||
QY 61 CTCGAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGACGAGAAATGGC 120
DB |||||
QY 121 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAGACAACTGTGTGCTTCAGAAAC 180
DB |||||
QY 181 CCTCAGCAGTTATGGATTTTTCAGAGGAATCCGAAATGACTTTCTGGAATCTTG 240
DB |||||
QY 241 AGCGCAGCTTTGGCACTAAAAGGGTCCACAACATTTCTACAATGAATACATCAGC 300
DB |||||
QY 301 CACCGAGAGCATCCACATGAACCTGACCCAGTGGGAGACACTGACCGACTTTTACCAG 360
DB |||||
QY 361 TGGCTGGGAGAGAGGCTTGTAAAGTACAGTGCACTGAAGCTGCTGGGAGCGCA 420
DB |||||
QY 421 GCATCCGGGAAACGGAAGAGTCTTACAGAGCTCCGCCAGCTCGCAAGGAAGAGAG 480
DB |||||
QY 481 TCGGCTCGATGAGATCATGTGAGCTCGAAGAGGAAAGAAAGGACCGACGACAGAC 540
DB |||||
QY 541 GCCTGGTTACGCGGGGATCGTTGAAATTTATACGAAGCTTCGGGAGAAATAT 600
DB |||||
QY 601 CACAAGAAAGGGTCTGTTAAGGAAGTGAATGACAGGTACACAGCTGTGGTAAAGATG 660
DB |||||
QY 661 ACTGACTTGGAGACAGGTGAAATGGAACGACTCATTTAGACAGACTCATTCGGCC 720
DB |||||
QY 721 CCGGGAAAGGGTCTAGTTTAAATGGGCTACAGAGGAATGAAGGCACTCTCGAA 780
DB |||||
QY 781 TCCATCAATGAGAAGGCTTTTCAGCCAGATAGTCAATGAACTGGACCTTTGAAAGGA 840
DB |||||
QY 841 CGCAGAGTTGAAGGTATTCAATATGAAGACATATCTAAACTTGGCTTGAATAATTT 900
DB |||||
QY 901 GATACAAACACATTGAAGCTGTGAAGCATCAATTTGCTGTAGCCAAAGCAGCTGTAC 960
DB |||||

QY 961 TCTACTGTGTAGGGATTTGTTTATTAATAAAAAAATCATCTATTAAATACT 1020
DB |||||
QY 1021 AGTGAATAGTTGGGTAAATTTATATAAAATCTATGTTTTTAAAGTGTAATAAAAAA 1080
DB |||||
QY 1081 AAAAAAATTTTAAAGTGTAATAAAAAA 1102
DB |||||

RESULT 2
AA85570
ID AA85570 standard; cDNA; 1390 BP.

XX AA85570;
XX DT 07-SEP-1999 (first entry)
XX cDNA of a gene coding for the murine kin17 protein.
DE Mouse; kin17 protein; cell proliferation; fertility;
XX hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme; ss.
OS Mus sp.
XX FR2772046-A1.
XX PD 11-JUN-1999.
XX PF 09-DEC-1997; 97FR-00015536.
XX PR 09-DEC-1997; 97FR-00015536.
XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
XX New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.
XX Claim 21; Page 35-36; 69pp; French.

CC The present sequence encodes a murine kin17 protein. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation

SQ Sequence 1390 BP; 482 A; 256 C; 353 G; 299 T; 0 U; 0 Other;

Query Match 72.5%; Score 799.2; DB 2; Length 1390;
Best Local Similarity 79.1%; Pred. No. 1.4e-148;
Matches 1099; Conservative 0; Mismatches 3; Indels 288; Gaps 1;

QY 1 ATGGCCAGTCCGATTTCTGAGCCCAAGGCATCGCAATAGATTAAGTCCAAAGG 60
DB |||||
QY 61 CTCGAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGACGAGAAATGGC 120
DB |||||


```
QY 121 TTTAAGTGTCTATGATCTGATCAAAAGACAACCTGTTGCTGCTTCAGAAAC 180
DB 121 TTTAAGTGTCTATGATCTGATCAAAAGACAACCTGTTGCTGCTTCAGAAAC 180
QY 181 CCTCAGCAGTTATGATTAATTTTTCAGAGGAATCCGAAATGACATTTCTGGAACCTTCG 240
DB 181 CCTCAGCAGTTATGATTAATTTTTCAGAGGAATCCGAAATGACATTTCTGGAACCTTCG 240
QY 241 AGCGCAGCTTTGGCACTAAAGGGTCCACAACAATTGTTCTCAATGAATACATCAGC 300
DB 241 AGCGCAGCTTTGGCACTAAAGGGTCCACAACAATTGTTCTCAATGAATACATCAGC 300
QY 301 CACCGAGAGCAGATCCACATGACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAAG 360
DB 301 CACCGAGAGCAGATCCACATGACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAAG 360
QY 361 TGGCTGGGAGAGAGGGCTTGTGTAAA----- 387
DB 361 TGGCTGGGAGAGAGGGCTTGTGTAAA----- 387
QY 388 ----- 387
DB 421 TACATAGACAGAGCCAGAAACCATCCGTCGGCACTGGAAATTAGAAAAAAGAAAG 480
QY 388 ----- 387
DB 481 CAAGATCTGCGATGAAGAAAAAACTGCCAAGTTTCATTGAGGAGCAGGTGAGAAAGGC 540
QY 388 ----- 387
DB 541 CTGGAAGGGAAGAGCAGGAGACACCTGTTTTCAGACACTTAGCCGAGAAATGAGGA 600
QY 388 ----- 387
DB 601 GAAAAAGTTACGTTCAATCTGAATAAGAGCGGGTGGCTCAGCGGAGCTACAACATCC 660
QY 388 -----GATACAGTGCATCAAGCTGCTGGGAGCGCAGCATCCGGGAAA 432
DB 661 AAGTCAAGCTCTTTGGGACCAAGTGCACTGAAGCTGCTGGGAGCGCAGCATCCGGGAAA 720
QY 433 CGGAAGAGTCTTCACAGAGCTCCGCCAGCTCGCGAAGAGAAAGTCCGGCCTGGAT 492
DB 721 CGGAAGAGTCTTCACAGAGCTCCGCCAGCTCGCGAAGAGAAAGTCCGGCCTGGAT 780
QY 493 GAGATCATGAGCTCGAAGAGGAAAAAGAACCGCACGAGCAGCGCTGGTTACAG 552
DB 781 GAGATCATGAGCTCGAAGAGGAAAAAGAACCGCACGAGCAGCGCTGGTTACAG 840
QY 553 CCGGGGATCGTTGTGAAAAATTATACGAAGAGCTTGGGAGAAATATCAAGAAAGAA 612
DB 841 CCGGGGATCGTTGTGAAAAATTATACGAAGAGCTTGGGAGAAATATCAAGAAAGAA 900
QY 613 GGGTCTGTAAGGAAGTATTGACAGGTACACAGCTGTGTTAAAGATGACTCTCGGA 672
DB 901 GGGTCTGTAAGGAAGTATTGACAGGTACACAGCTGTGTTAAAGATGACTCTCGGA 960
QY 673 GACAGGCTGAAATCTGGACACGACTCATTTAGAGACAGTCAATTCGGCCCCCGGGGAAAGG 732
DB 961 GACAGGCTGAAATCTGGACACGACTCATTTAGAGACAGTCAATTCGGCCCCCGGGGAAAGG 1020
QY 733 GTTCTAGTTTAAATGAGGGCTACAGAGGAATGAAGGCACTCTCGAATCCATCAATGAG 792
DB 1021 GTTCTAGTTTAAATGAGGGCTACAGAGGAATGAAGGCACTCTCGAATCCATCAATGAG 1080
QY 793 AAGGCTTTTTCAGCCAGATAGTCAATGAACCTTGAAGAGCGCAGAGTTGAA 852
DB 1081 AAGGCTTTTTCAGCCAGATAGTCAATGAACCTTGAAGAGCGCAGAGTTGAA 1140
QY 853 GGTATTCAATATGAAGACATATCTAAACTTCTGAGTTTGAATTTGATAACAACACA 912
DB 1141 GGTATTCAATATGAAGACATATCTAAACTTCTGAGTTTGAATTTGATAACAACACA 1200
```

```
QY 913 TTGAAACTGTGAGCATCAAAATGGTGTGTAGCCAGGCACTGTGTAACCTCTACTGTGTTA 972
DB 1201 TTGAAACTGTGAGCATCAAAATGGTGTGTAGCCAGGCACTGTGTAACCTCTACTGTGTTA 1260
QY 973 GGGGATTTGTTTGTATTAAAAAATAAATCATCTATTTAAATACTAGTGAATAGTTG 1032
DB 1261 GGGGATTTGTTTGTATTAAAAAATAAATCATCTATTTAAATACTAGTGAATAGTTG 1320
QY 1033 GGTAAATTTATATAAATCTATGTTTTTTTAAAGTGTAAAAAATAAATAAATAAATAA 1092
DB 1321 GGTAAATTTATATAAATCTATGTTTTTTTAAAGTGTAAAAAATAAATAAATAAATAA 1380
QY 1093 AAAAAAATAA 1102
DB 1381 AAAAAAATAA 1390
RESULT 3
AAQ79936
ID AAQ79936 standard; cDNA; 1458 BP.
XX
AC AAQ79936;
XX
DT 25-MAR-2003 (revised)
DT 06-SEP-1995 (first entry)
XX
DE Murine Kin17 cDNA.
XX
KW chromosomal rearrangement; kin17 protein; SOS DNA repair system; RecA;
KW genotoxic agent; zinc finger; DNA binding protein; ss.
XX
OS Mus musculus.
XX
FH Key
FT primer_bind Location/Qualifiers
FT /tag= b complement(1..21)
FT /note= "Oligo L (AAQ79946) binds to complementary strand"
FT misc_feature 22..1434
FT /tag= 1 /label= kin17_cDNA
FT /note= "nucleotides 1-1414; the genomic DNA contains at
FT /note= "least two introns within this sequence, see Comments"
FT primer_bind 32..49
FT /tag= c
FT /note= "Oligo C (AAQ79938) binding site"
FT CDS 46..1221
FT /tag= a
FT /product= "kin17"
FT /note= "N's in the sequence denote illegible residues"
FT primer_bind complement(67..86)
FT /tag= d
FT /note= "Oligo S (AAQ79947) binds to complementary strand"
FT primer_bind 274..297
FT /tag= e
FT /note= "Oligo D (AAQ79939) binding site"
FT primer_bind complement(339..360)
FT /tag= f
FT /note= "Oligo K (AAQ79945) binds to complementary strand"
FT primer_bind 451..474
FT /tag= g
FT /note= "Oligo J (AAQ79944) binding site"
FT primer_bind complement(550..567)
FT /tag= h
FT /note= "Oligo E (AAQ79940) binds to complementary strand"
FT primer_bind 802..825
FT /tag= i
FT /note= "Oligo F (AAQ79941) binding site"
FT primer_bind complement(839..862)
FT /tag= j
FT /note= "Oligo G (AAQ79942) binds to complementary strand"
FT primer_bind complement(1435..1458)
FT /tag= k
FT /note= "Oligo B (AAQ79937) binds to complementary strand"
```

XX FR2706487-A1.
 XX PD 23-DEC-1994.
 XX PF 15-JUN-1993; 93PR-00007171.
 XX PR 15-JUN-1993; 93PR-00007171.
 XX PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX PI Angulo-Mora JF, Tissier A, Prelat G, Mauffrey P, Guilly M;
 XX DR WPI; 1995-039031/06.
 XX PT Purified murine kin17 protein prepn. for detecting chromosomal
 PT rearrangements - also related antibodies, human and murine DNA, primers,
 PT probes and vectors, used to assess damage caused by genotoxic agents.
 XX PS Claim 9; Page 33; 54pp; French.
 XX CC The murine Kin17 protein includes a zinc finger domain (see AAR66766),
 CC recognises single- and double-stranded DNA (partic. regions of secondary
 CC structure), has apparent mol. wt. 43 kD and is recognised by both anti-
 CC kin17 antibodies and antibodies against the RecA protein of E.coli. The
 CC Kin17 protein is involved in DNA repair; it can be used to monitor
 CC chromosomal rearrangements following exposure to genotoxic agents. The
 CC kin17 cDNA sequence AAQ79936 consists of a 1414 nucleotide sequence,
 CC flanked by primer binding sites; the genomic kin17 DNA is claimed in
 CC which an intron is inserted at position 137 of the 1414 nucleotide cDNA
 CC and a second intron is located between nucleotides 339-429 of the cDNA.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX SQ Sequence 1458 BP; 496 A; 272 C; 372 G; 311 T; 0 U; 7 Other;

Query Match 71.7%; Score 790; DB 2; Length 1458;
 Best Local Similarity 78.5%; Pred. No. 9.1e-147;
 Matches 1091; Conservative 0; Mismatches 11; Indels 288; Gaps 1;

QY 1 ATGGGCAAGTCGGATTTCTGAGCCCAAGGCCATCGCCAAATAGAAATTAAGTCCAAAGG 60
 DB 46 ATGGGCAANTCGGATTTCTGAGCCCAAGGCCATCGCCAAATAGAAATTAAGTCCAAAGG 105
 QY 61 CTCGAGAGCTTCGGTGTCTGACGATGTCGCAAGAGCAATGCGGACGAGAAATGCG 120
 DB 106 CTCGAGAGCTTCGGTGTCTGACGATGTCGCAAGAGCAATGCGGACGAGAAATGCG 165
 QY 121 TTTAAGTGTCACTGTATGTCGAATCTCATCAAGACAACTGTGCTGCTTCAGAAAC 180
 DB 166 TTTAAGTGTCACTGTATGTCGAATCTCATCAAGACAACTGTGCTGCTTCAGAAAC 225
 QY 181 COTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTTCTGGAATCTCTG 240
 DB 226 COTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTTCTGGAATCTCTG 285
 QY 241 AGGCGACGTTTGGCAGCTTAAAGGGTCCCAACACATTTGCTACAAGATACATCAGC 300
 DB 286 AGGCGACGTTTGGCAGCTTAAAGGGTCCCAACACATTTGCTACAAGATACATCAGC 345
 QY 301 CACCCGAGAGCAGCATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAG 360
 DB 346 CACCCGAGAGCAGCATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAG 405
 QY 361 TGGCTGGGCGAGAGGGCTTTGTGTAAA----- 387
 DB 406 TGGCTGGGCGAGAGGGCTTTGTGTAAAGTGGATGAGACACCGAAGGCTGGTACATTACG 465
 QY 388 ----- 387
 DB 466 TACATAGACAGAGACCCAGAAACCATCCGTCGGCAACTGGAATTAGAAAAAAGAGAAG 525
 QY 388 ----- 387

DB 526 CAAGATCTGGAGATGAAGAAAAAAGTCCCAAGTTTCATTGAGGAGCAGGTCAAGAGAGCG 585
 QY 388 ----- 387
 DB 586 CTNGAAGGGAAGAGAGAGGAGACACCTGTTTTTACAGAACTTAGCCGAGAAAAATGAGGAA 645
 QY 388 ----- 387
 DB 646 GAAAAAGTTACGTTCAATCTGAAATAAAGAGCGGGTGGCTCAGCGGGAGACTACAACATCC 705
 QY 388 -----GGTACCAGTGCCTGAAGCTGCTGGGGAGCGCAGCATCCCGGAAA 432
 DB 706 AAGTCAAGCTCTTTGGGACCAAGTGCATGAAGCTGCTGGGGAGCGCAGCATCCCGGAAA 765
 QY 433 CGGAAAGAGTCTTTCACAGAGCTCCGCCAGCTCGGAGGAGAAAGAAAGTTCGGCCCTCGAT 492
 DB 766 CGGAAAGAGTCTTTCACAGAGCTCCGCCAGCTCGGAGGAGAAAGAAAGTTCGGCCCTCGAT 825
 QY 493 GAGATCATGGAGCTCGAAGAGGAAAGAAAGGACCGCACGACAGAGCGCTGGTTACAG 552
 DB 826 GAGATCATGGAGCTCGAAGAGGAAAGAAAGGACCGCACGACAGAGCGCTGGTTACAG 885
 QY 553 CCGGGATCGTTGTGAAAATTTATAACGAAAGAGCTTGGGGAGAAATATACAAAGAGAAA 612
 DB 886 CCGGGATCGTTGTGAAAATTTATAACGAAAGAGCTTGGGGAGAAATATACAAAGAGAAA 945
 QY 613 GGGGTCGTTAAGGAGTGAATGACAGGTACACAGCTGTGTGTAAGATGACTGACTCTGGA 672
 DB 946 GGGGTCGTTAAGGAGTGAATGACAGGTACACAGCTGTGTGTAAGATGACTGACTCTGGA 1005
 QY 673 GACAGGCTGAAACTGGACCGACAGCTCATTTAGAGACAGTCATTCGGGCCCGGGGAAAAGG 732
 DB 1006 GACAGGCTGAAACTGGACCGACAGCTCATTTAGAGACAGTCATTCGGGCCCGGGGAAAAGG 1065
 QY 733 GTTCTAGTTTTAAATGGAGGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCAATGAG 792
 DB 1066 GTTCTAGTTTTAAATGGAGGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCAATGAG 1125
 QY 793 AAGGCTTTTTCAGGCACGATAGTCAATTGAAACTGGAACCTTTGAAAGGACGCGAGATTGAA 852
 DB 1126 AAGGCTTTTTCAGGCACGATAGTCAATTGAAACTGGAACCTTTGAAAGGACGCGAGATTGAA 1185
 QY 853 GGTATTCATATGAAGACATATCTAAACTCTGCTGAGTTTGAABAATTTGATAACAACACA 912
 DB 1186 GGTATTCATATGAAGACATATCTAAACTCTGCTGAGTTTGAABAATTTGATAACAACACA 1245
 QY 913 TTGAACTGTGAAGCATCAAAATTTGGTGTAGCCAGGCACTGTGTAACTCTACTGTGTTA 972
 DB 1246 TTGAACTGTGAAGCATCAAAATTTGGTGTAGCCAGGCACTGTGTAACTCTACTGTGTTA 1305
 QY 973 GGGGATTTGTTTGTATTTAAAAAATAAATCATCTATTATTAATACTAGTGAATAGTTG 1032
 DB 1306 GGGGATTTGTTTGTATTTAAAAAATAAATCATCTATTATTAATACTAGTGAATAGTTG 1365
 QY 1033 GGTAAATTTATAAATAAATCTATGTTTTTTTTTAAGTGTAAAAAATAAATAAATAAATAA 1092
 DB 1366 GGTAAATTTATAAATAAATCTATGTTTTTTTTTAAGTGTAAAAAATAAATAAATAAATAA 1425
 QY 1093 AAAAAAATAA 1102
 DB 1426 AAAAAAATAA 1435

RESULT 4
 AAX85551
 ID AAX85551 standard; cDNA; 1002 BP.
 XX AAX85551;
 AC AAX85551;
 XX 07-SEP-1999 (first entry)
 XX cDNA of a gene coding for a truncated human kin17 protein.
 DE
 XX

KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme; ss.
 XX
 OS Homo sapiens.
 XX
 PN FR2772046-A1.
 XX
 PD 11-JUN-1999.
 XX
 XX 09-DEC-1997; 97PR-00015536.
 XX
 XX 09-DEC-1997; 97PR-00015536.
 XX
 XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX
 XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 XX WPI; 1999-359999/31.
 XX
 XX New DNA coding for human kin17 protein - useful for controlling cell
 XX proliferation or fertility.
 XX
 XX Claim 5; Page 32; 69pp; French.
 XX
 XX The present sequence encodes a truncated human kin17 protein with amino
 XX acids 129-228 deleted. The mammalian kin17 protein is useful for
 XX preparing a medicament for controlling cell proliferation or for
 XX controlling fertility. The medicaments can also be used to treat
 XX hyperproliferative diseases. Fragments between amino acids 55 and 235
 XX (preferably between amino acids 129 and 228) of a mammalian kin17 protein
 XX are useful for regulating the interaction between proteins and curved
 XX DNA. The fragment can be used to block replication of HIV or its
 XX integration into the human genome or to target repair enzymes to curved
 XX DNA sites. Expression vectors for kin17 can be used for controlling cell
 XX proliferation
 XX
 XX Sequence 1002 BP; 343 A; 189 C; 230 G; 240 T; 0 U; 0 Other;
 XX
 XX Query Match 63.9%; Score 704.6; DB 2; Length 1002;
 XX Best Local Similarity 86.1%; Pred. No. 6.8e-130;
 XX Matches 819; Conservative 0; Mismatches 119; Indels 13; Gaps 3;
 QY 1 ATGGGCAAGTCGGATTTCTGAGCCCAAGCCATCCCAATAGATTAGTCCAAAGGG 60
 DB 58 ATGGGGAAGTCGGATTTCTTACTCCCAAGCTATCCCAACAGGATCAAGTCCAAAGGG 117
 QY 61 CTCCAGAGCTTCGCTGTACTGCGATGTCGCAAGCAATGCCGCGACGAGAATGGC 120
 DB 118 CTGCAAGAGCTACCGCTGTATGTCAGATGTCGCAAGCAATGCCGCGACGAGAATGGC 177
 QY 121 TTTAAGTGTCATCTGTATGTCTGAATCTCATCAAGACAACTGTGTGGCTTCAGAAAC 180
 DB 178 TTTAAGTGTCATCTGTATGTCGCAATCTCATCAAGACAACTGTGTGGCTTCAGAAAC 237
 QY 181 CCTCAGCAGTTATGGATTTATTTTTCAGAGAAATCCGAATGACTTTCTGGAATCTCTG 240
 DB 238 CCTCAGCAGTTATGGATTTATTTTTCAGAGAAATCCGAATGACTTTCTGGAATCTCTC 297
 QY 241 AGGCGAGCTTTGGCAGCTTAAAGGGTCCACAAACAACTGTCTCAATGAATACATCAGC 300
 DB 298 AGGAGAGCTTTGGCAGCTTAAAGGGTCCACAAACAACTGTCTCAATGAATACATCAGC 357
 QY 301 CACGAGAGCAGCAGTCCATCAATGAACGCTACCCAGTGGGAGACACTGACCGACTTTACCAAG 360
 DB 358 CACGAGAGCAGCAGTCCATCAATGAACGCTACCCAGTGGGAGAACTCTGACTGATTTTAAAG 417
 QY 361 TGGCTGGGCGAGAGGGCTTGTAAAGGTACAGTGCAGTGAAGCTGCTGGGAGCGCA 420
 DB 418 TGGCTGGGCGAGAGGGCTTGTGCAAA-----AGTGCAGTGAAGAGCATAGGAAGTTCA 471
 QY 421 GCATCCGGGAACGGAAGAGTCTTCAGAGAGTCCGCGCAGCT-----GCCAAGAG 474

DB 472 GCATCAGTGAACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAA 531
 QY 475 AAGAGCTCGCCCTGGATCAGATCATGGAGCTCGAAGAGAGAAAGAAAGGACCGCACGG 534
 DB 532 AAGAAATCTGCACTGGATGAATCATGGAGATTGAAGAGGAAAGAAAGAACTGCCCCG 591
 QY 535 ACAGAGCCTCGTTACAGCGCGGGATCGTTGTGAAAAATTATAACGAAGAAAGCTTGGGGAG 594
 DB 592 ACAGACTACTGGCTACAGCTGAAATTTATGTGAAAAATTATAACGAAGAACTGGGAGAG 651
 QY 595 AATATCAAGAGAAAGGGGTCGTTAAGGAAGTCAATGACAGGTACACAGCTGTGGTA 654
 DB 652 AATATCATAAGAAAGAAAGGCTATTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTG 711
 QY 655 AAGATCACTGACTCTCGAGACAGGCTGAAACTGACAGACTCATTTAGAGACAGTCAAT 714
 DB 712 AAGATGATTGATCTCGAGACAGGCTGAAACTGACAGACTCATTTAGAGACAGTAAAT 771
 QY 715 CCGGCCCCGGGAAAGGGTTCTAGTTTAAATGGAGGCTACAGAGAAATGAAGGCACT 774
 DB 772 CCAGCACCAGGAAAGAAAGTCTTAGTTTAAATGGAGGCTACAGAGAAATGAAGGTACC 831
 QY 775 CTCGAATCCATCAATGAGAGGCTTTTTCAGCCAGCATAGTCAATGAAACTGGACCTTTG 834
 DB 832 CTAGAATCCATCAATGAGAGGCTTTTTCAGCTACTATCGTCAATGAAACTGGCCCTTTA 891
 QY 835 AAAGGACGAGAGCTTGAAGGTATTCAATATGAAGACATATCTAAACTTGTCTGAGTTTGA 894
 DB 892 AAAGGACGAGAGTGAAGGAATCAATATGAAGACATTTCTAAACTTGTCTGAGTTTGA 951
 QY 895 AAATTTGATAACCAACATTTGAAACT-GTGAAGCATCAAAATTTGGTGTAGC 944
 DB 952 AAATTTGATAACATACCTTTAAATCTTAAAGCATCAAAATTTGGTGTAGC 1002
 RESULT 5
 AAX85549
 ID AAX85549 standard; cDNA; 1296 BP.
 AC AAX85549;
 XX 07-SEP-1999 (first entry)
 DT
 XX cDNA of a gene coding for the human kin17 protein.
 DE
 XX Human; kin17 protein; cell proliferation; fertility;
 KW hyperproliferative disease; protein interaction; curved DNA;
 KW HIV replication; HIV integration; repair enzyme; ss.
 XX
 OS Homo sapiens.
 XX
 XX FR2772046-A1.
 XX
 XX 11-JUN-1999.
 XX
 XX 09-DEC-1997; 97PR-00015536.
 XX
 XX 09-DEC-1997; 97PR-00015536.
 XX
 XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX
 XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 XX WPI; 1999-359999/31.
 XX
 XX New DNA coding for human kin17 protein - useful for controlling cell
 XX proliferation or fertility.
 XX
 XX Claim 1; Page 30-31; 69pp; French.
 XX
 XX The present sequence encodes a human kin17 protein. The mammalian kin17
 XX protein is useful for preparing a medicament for controlling cell
 XX proliferation or for controlling fertility. The medicaments can also be

CC used to treat hyperproliferative diseases. Fragments between amino acids
CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
CC kin17 protein are useful for regulating the interaction between proteins
CC and curved DNA. The fragment can be used to block replication of HIV or
CC its integration into the human genome or to target repair enzymes to
CC curved DNA sites. Expression vectors for kin17 can be used for
CC controlling cell proliferation
XX
SQ Sequence 1296 BP; 453 A; 240 C; 310 G; 293 T; 0 U; 0 Other;

Query Match 38.3%; Score 422.2; DB 2; Length 1296;
Best Local Similarity 66.3%; Pred. No. 5.6e-74;
Matches 821; Conservative 0; Mismatches 123; Indels 295; Gaps 3;

QY 1 ATGGGCAAGTCGGATTTCTGAGCCCAAGGCCATCGCCAAATGAAATTAAGTCCAAAGGG 60
DB |||||
QY 58 ATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATGCCCAACAGGATCAAGTCCAAAGGG 117
DB |||||
QY 61 CTCGAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAGATGGC 120
DB |||||
QY 118 CTGCGAAGCTACGCTGGTATTGCCAGATGTGCCAGACAGTGGCCGGACGAGATGGC 177
DB |||||
QY 121 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAGACAACTGTGCTGGCTTCAGAAAAC 180
DB |||||
QY 178 TTTAAGTGTCACTGTATGTCTGAATCTCATCAAGACAACTGTGCTGGCTTCAGAAAAC 237
DB |||||
QY 181 CCTCAGCAGTTATGATTAATTTTTCAGAGGAATTCGGAATGACCTTTCGGAACCTTCG 240
DB |||||
QY 238 CCTCAGCAGTTATGATTAATTTTTCAGAGGAATTCGGAATGACCTTTCGGAACCTTC 297
DB |||||
QY 241 AGCGCAGCTTTGGCAGTAAAGGGTCCACACAACTGTCTACAAATGATACATCAGC 300
DB |||||
QY 298 AGGACGCTTTGGCAGTAAAGGGTCCACAACTGTCTACAAATGATACATCAGC 357
DB |||||
QY 301 CACCGAGAGCATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACCTTTTACCAAG 360
DB |||||
QY 358 CACCGAGAGCATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACCTTTTACCAAG 417
DB |||||
QY 361 TGGCTGGGAGAGAGGGCTTTGTGTAAC----- 387
DB |||||
QY 418 TGGCTGGGAGAGAGGGCTTTGTGTAACGGTGGAGACACCAAGGGCTGGTATATTCAG 477
DB |||||
QY 388 ----- 387
DB 478 TACATAGACAGGAGCCAGAAACTATCCCGCGGCAACTGGAACTGGAGAAAAGAAAG 537
QY 388 ----- 387
DB 538 CAGGACCTTGATGATGAAGAAAACCTGCCAAATTTATTGAAGAGCAAGTGAGAGAGGC 597
QY 388 ----- 387
DB 598 CTGGAAGGAAGGAACAGGAGGTCCCTACTTTTACGGAATTAAGCAGAGAAAATGATGAA 657
QY 388 ----- 387
DB 658 GAGAAAGTCACGTTTAATTTGAGTAAAGGAGCATGTAGCTATCCGGAGCAACATCTTCC 717
QY 388 -----GGTACCAGTGCACCTGAAGCTGCTGGGAGCGCAGCATCCGGGAAA 432
DB |||||
QY 718 AAGTCAAGTACTCTGGGACCGAGTGCCTGAGACGATAGGAAGTTTCAGCATCAGTGA 777
DB |||||
QY 433 CGGAAAGAGTCTTTCAGAGAGCTCCGCCAGCCTGC-----GAAGAAGAAAGTCCGCC 486
DB |||||
QY 778 CGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAGAAAAGAAAAGAAATCTGCA 837
QY 487 CTGGATGAGATCATGAGCTGAGAGGAAAGAAAAGACCGCAGCAGAGCGCTGG 546
DB |||||
QY 838 CTGGATGAAATCATGGAGATTGAAGAGGAAAAGAAAAGAACTGCCGAGCAGACTACG 897
DB |||||
QY 547 TTACAGCCGGGGATCGTTGTGAAAATTATAACGAAGAGCTTGGGAGAGAAATATCAAG 606
DB |||||
QY 898 CTACAGCTTGAATTTATTGTGAATTTATTAACCAAGAACTGGGAGAGAAATATCATAG 957
DB |||||

QY 607 AAGAAAGGGTCTGTTAAAGAAAGTATTGACAGGTACACAGCTGTGGTAAAGATGACTGAC 666
DB |||||
QY 958 AAAAAGGCTATTGTTAAGAAAGTAAATGACAAATATACAGCTGTGTGAAAGATGATTGAT 1017
DB |||||
QY 667 TCTGAGAGCAGGCTGAAACTGGACAGACTCATTTAGAGACAGTCAATTCGGGCCCGGGG 726
DB |||||
QY 1018 TCTGGAGACAGCTGAAACTTTGACAGACTCATTTAGAGACAGTAATTTCCAGACCAGGA 1077
DB |||||
QY 727 AAAAGGGTCTAGTATTTTAAATGAGGGCTACAGAGAAATGAAGGCACTCTCGAATCCATC 786
DB |||||
QY 1078 AAAAGAATTTCTAGTTTTAAATGAGGGCTACAGAGAAATGAAGGTACCTAGAAATCCATC 1137
DB |||||
QY 787 AATCAGAGGCTTTTTCAGCCACGATAGTCAATGAACTGGACCTTTTGAAGAGCGCAGA 846
DB |||||
QY 1138 AATGAGAAAGACTTTTTCAGCTACTATCTGCTCAATGAACTGGCCCTTTTAAAGAGCGCAGA 1197
DB |||||
QY 847 GTTGAAGTATTTCAATATGAAGACATATCTAAACTTGTCTGAGTTTGAAGATTTGATAAC 906
DB |||||
QY 1198 GTTGAAGAAATTTCAATATGAAGACATTTCTAACTTGTCTGAGTTTGAAGATTTGTTAAC 1257
DB |||||
QY 907 AACACATTTGAAACT-GTGAAGCATCAAAATTTGGTGTAGC 944
DB |||||
QY 1258 AATACCTTTAAATCTTAAAGCATCAAAATTTGGTGTTCGC 1296
DB |||||

RESULT 6
ADQ56943
ID ADQ56943 standard; DNA; 679 BP.
XX
AC ADQ56943;
XX
DT 21-OCT-2004 (first entry)
XX
DE Novel canine microarray-related DNA sequence SeqID8245.
XX
KW canine microarray; drug screening; toxicity assay;
KW environmental pollutant; cellular response; gene expression profile;
KW toxic response; liver necrosis; fatty liver disease;
KW protein adduct formation; hepatitis; dogs; de.
XX
OS Canis familiaris.
XX
PN WO2004063324-A2.
XX
PD 29-JUL-2004.
XX
PF 05-MAY-2003; 2003WO-US013853.
XX
PR 03-MAY-2002; 2002US-0377240P.
XX
PA (GENE-) GENE LOGIC INC.
PA (PFIZ) PFIZER PROD INC.
XX
PI Diggins JC, Porter M, Wei T;
XX
XX WPI; 2004-561890/54.
DR
XX
PT New isolated nucleic acid molecule, useful for drug screening and
PT toxicity assays or for assessing the impact, including toxicity, of a
PT compound, pharmaceutical agent or environmental pollutant on a cell or
PT living organism.
XX
XX Claim 1; SEQ ID NO 8245; 41pp; English.
XX
CC This invention is related to a novel isolated canine nucleic acid
CC sequences and the construction of canine microarrays containing a
CC significant portion of the canine genome. The isolated canine nucleic
CC acid sequences of the invention may be useful for drug screening and
CC toxicity assays. The invention is therefore useful for assessing the
CC impact, including toxicity, of a compound, pharmaceutical agent or
CC environmental pollutant on a cell or living organism. The methods are
CC useful for detecting genes that are up- or down-regulated in canines in a

CC disease state. The sequences are useful as diagnostic agents or markers
CC to detect a cellular response in a sample individually or as part of a
CC gene expression profile. It is also useful as a target for agents that
CC modulate gene expression or activity. The database is useful for
CC producing electronic Northern blots that allow the user to determine the cell
CC type or tissue in which a given gene is expressed and to allow
CC determination of the abundance or expression level of a given gene in a
CC particular tissue or cell. The methods are useful for determining the
CC similarity of a toxic response to one or more individual compounds. The
CC methods are useful for predicting at least one toxic response or the
CC likelihood that a compound or test agent will induce various specific
CC pathologies such as those of the liver (liver necrosis, fatty liver
CC disease, protein adduct formation or hepatitis), those of the kidney,
CC heart, brain or testes, or other pathologies associated with at least one
CC of the toxins. The methods are also useful for predicting or elucidating
CC the potential cellular pathways influenced, induced or modulated by the
CC compound or test agent due to the similarity of the expression profile
CC compared to the profile induced by a known toxin. The present sequence is
CC that of a canine DNA sequence which was claimed for use during the
CC production of a canine microarray of the invention.
XX
SQ Sequence 679 BP; 254 A; 106 C; 127 G; 185 T; 0 U; 7 Other;

Query Match 35.0%; Score 385.4; DB 13; Length 679;
Best Local Similarity 79.3%; Pred. No. 9.6e-67;
Matches 494; Conservative 0; Mismatches 117; Indels 12; Gaps 3;

QY 468 GAAGAAGAAGATCGGCCCTCGATGATCATGGAGCTCGAAGAGAGAAAGAGGAC 527
DB 24 GAAAAAAGAAATCTGCNNCGACGAATCATNAGATTGAAGAGGAAAAAGAAAGAC 83

QY 528 CGACGAGACAGCCCTGGTTACAGCCGGGATCGTTGTGAAATTTATACGAGAGCT 587
DB 84 TGCCCGAACAGACCTGGCTACAGCCGGGAATCATCGTGAATAATATACCAAAACT 143

QY 588 TGCGGAGAAATATCACAAGAAGAAAGGGTCTGTAAAGAGTGAATGACAGGTACACAGC 647
DB 144 TGAGAGAAATATCATAGAAAGGGTGTGTGTAAGAGTGAATGACAAATATACAGC 203

QY 648 TGTGTAAGATGACTCTCGAGACAGGCTGAACTGGACAGACTCATTTAGAGAC 707
DB 204 TGTGTAAGATGACTCTCGAGACAGCTGAACTTTGACAGACTCATTTAGAAAC 263

QY 708 AGTCATTTCCGGCCCGGGGAAAGGGTCTAGTTTAAATGGAGCTACAGAGGAATGA 767
DB 264 AGTAATTCAGCACAGGAAAGAAATCTTGTGTTTAAATGGAGGCTACAGAGAAATGA 323

QY 768 AGGCATCTCGAATCCATCAATGAGAGGCTTTTTCAGCCACGATAGTCAATGAAACTGG 827
DB 324 AGGCATCTTGAATCCATCAATGAGAGACTTTTCTCAGCTACATATCAATGAACTGG 383

QY 828 ACCTTTGAAAGGACGACAGATTTGAAGTATTCATATGAAGACATATCTAACTTGTCTG 887
DB 384 CCCTTTAAAGGACGACAGATTTGAAGAAATTCATACGAGACATTTCTAACTTGCCTG 443

QY 888 AGTTTGAATAATGATTAACAACACATTCGAACTGTGAAGCATCAATTTGTTGTAGCCAA 947
DB 444 AGTTTGAATAATCTTTT---AACACATTTAAATCCCTAAACATCAATTTGTTGTTCACCA 500

QY 948 GGCACTGTGAATCTACTGTGTTAGGGATTTCTGTTTGTATTAATAAAAAAATCAAT 1007
DB 501 GGCAATATGAGACTCTACTGTGTTAGGGTGTCTTTTGTAT-----AAGCAATATGAT 555

QY 1008 CTATTTAAATCTAGTGAATAGTTG---GGTAAATTTATATATAATCTATGTTTTTTT 1063
DB 556 TTATTTAAATCTAGTGAATAGTTGTTTTCAGCAAGCTTATAGCAAAATCTAATATATGCT 615

QY 1064 TAAGTGTAAAAAATAAAAAA 1086
DB 616 CACGAAGTATCAGAACATGTAA 638

RESULT 7

AAx85552
ID AAX85552 standard; cDNA; 1002 BP.
XX
AC AAX85552;
XX
DT 07-SEP-1999 (first entry)
XX
DE Probe derived from cDNA of a gene coding for the human kin17 protein.
XX
XX Human; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme; probe; ss.
XX
OS Homo sapiens.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX
XX WPI; 1999-359999/31.
DR
PT New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.
XX
PS Claim 7; Page 32-33; 69pp; French.
XX
XX The present sequence is derived from nucleotides 207-1208 of AAX85549,
CC and is a probe used to isolate human kin17 cDNA. The mammalian kin17
CC protein is useful for preparing a medicament for controlling cell
CC proliferation or for controlling fertility. The medicaments can also be
CC used to treat hyperproliferative diseases. Fragments between amino acids
CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
CC kin17 protein are useful for regulating the interaction between proteins
CC and curved DNA. The fragment can be used to block replication of HIV or
CC its integration into the human genome or to target repair enzymes to
CC curved DNA sites. Expression vectors for kin17 can be used for
CC controlling cell proliferation
XX
SQ Sequence 1002 BP; 374 A; 179 C; 235 G; 214 T; 0 U; 0 Other;

Query Match 29.9%; Score 329.2; DB 2; Length 1002;
Best Local Similarity 82.4%; Pred. No. 1.4e-55;
Matches 392; Conservative 0; Mismatches 78; Indels 6; Gaps 1;

QY 388 GGTACCAAGTGCATGAAGCTGTGGGAGGCGAGATCCGGGAACCGAAGAGCTCTTCA 447
DB 527 GGACCGAGTGCATGAAGACGATAGGAAGTTCAGCATCAGTGAACGAAAGAAATCTTCC 586

QY 448 CAGAGCTCCGCCAGCTGC-----GAAGAAGAAGTCCGCCCTGGATGAGATCATG 501
DB 587 CAGAGCTCAACTCAGTCTAAAGAAAAGAAAAGAAAATCTGCACCTGATGAATCATG 646

QY 502 GAGCTCGAAGAGGAAAAAGAACCGACGACAGCCTGTTTACAGCCGGGATC 561
DB 647 GAGATTGAAGGGAAGAAAAGAAAGAACTGCCGACAGACTACTGCTACAGCTGAAAT 706

QY 562 GTTGTGAAAAATTATACGAAGAGCTTTGGGAGAGAAATATCAAGAAGAAAGGGTCTT 621
DB 707 ATTGTGAAAAATTATACCAAGAACTCGGAGAGAAATATCATAGAAAAGGGCTATTGTT 766

QY 622 AAGCAAGTATTGACAGGTACACAGCTGTGTGAAGATGACTGCTCTGGAGACAGGCTG 681
DB 767 AAGCAAGTAAATTTGACAAATATATACAGCTGTGTGTGAAGATGATTGATCTCGAGAC 826

QY 682 AACTGACCAAGACTCATTTTAGACAGACTCATTCGGGCCCCGGGAAAAAGGGTTCTAGTT 741

Db 827 AAACITGACCACTCATTTAGAGCAGTAATTCAGACACGAGGAAAAAGAAATCTAGTT 886
QY 742 TTAATGAGGCTACAGAGAAATGAAGCACTCTCGAAATCCATCAATGAGAAGCTTTT 801
Db 887 TTAATGAGGCTACAGAGAAATGAAGTACCTAGATCCATCAATGAGAAGACTTTT 946
QY 802 TCAGCCACATAGTCATTTGAACTGACCTTTGAAAGGACGAGATTGAAAGTAT 857
Db 947 TCAGCTACTATCGTCATTTGAAATCGCCCTTTAAAGAGGACGAGAGTTGAAGGAAT 1002

RESULT 8
AAH34095
ID AAH34095 standard; cDNA; 591 BP.
XX
AC AAH34095;
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1177.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; es.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
PD
PD 05-APR-2001.
PF 28-SEP-2000; 2000WO-US026524.
XX
XX 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI; 2001-235357/24.
DR P-PSDB; AAG74690.
DR
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
PS
PS Claim 1; Page 2977-2978; 9803pp; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922

XX Sequence 591 BP; 199 A; 95 C; 114 G; 182 T; 0 U; 1 Other;
Query Match 24.4%; Score 269.2; DB 4; Length 591;
Best Local Similarity 77.8%; Pred. No. 9.4e-44;
Matches 325; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 646 GCTGTGTTAAAGATGACTGACTCTCGAGACAGGCTGAAACTGGACAGACTCATTTAGAG 705
Db 28 GCTGAGGAATTCGGCACGAGTCTGGAGACAAGCTGAAACTTGACAGACTCATTTAGAG 87
QY 706 ACAGTCATTTCGGCCCCGGGAAAAAGGGTTCTAGTTTTTAAATGGAGGCTACAGAGGAAAT 765
Db 88 ACAGTAAATTCAGCACACAGGAAAAAGAAATTCAGTTTTTAAATGGAGGCTACAGAGGAAAT 147
QY 766 GAAGGCACTCTCGAATCCATCAATGAGAGGCTTTTTCAGCCACGATAGTCATTTGAAACT 825
Db 148 GAAGGTACCTAGATCCATCAATGAGAGACTTTTTCAGCTACTATCGTCTATTGAAACT 207
QY 826 GGACCTTTGAAAGGACGAGAGTTGCAAGGTATTCATATGAAGACATATCTAAACTTGCT 885
Db 208 GGCCCTTTAAAGAGGACGAGAGTTGAAGAAATTCATATGAAGACATTTCTAAACTTGCC 267
QY 886 TGAGTTTGAATAATTTGATAACAACACATTGAAACTGTGAAAGCATCAAAATTTGGTGTAGCC 945
Db 268 TGAGTTTGAATAATTTGTTAACCAATACATTAATACTTAAAGCATCAAAATTTGGTGTAGCC 327
QY 946 AAGGCACTGTGTACTCTACTCTGTTAGGGATTTTGTATTTAAAAAATAATCAATCAATCAAT 1005
Db 328 AAGGCATTATGAGACTCTACTCTGTTAGGGTATATTTCTTTGTATATAAACAACAGGTTT 387
QY 1006 ATCTATTTAAATACTAGTGAATAGTTGGTAAATTTTATAATAAATCTATGTTTTTTT 1063
Db 388 TTGAAATAATATTACTGTATAGTTAGTTGTTTCAGCTAAACTTTTGAGAGAAATTTAATAT 445

RESULT 9
ADP22451
ID ADP22451 standard; DNA; 1372 BP.
XX
AC ADP22451;
DT 12-AUG-2004 (first entry)
XX
DE Sea-squirt (Clona intestinalis) zinc finger protein coding sequence #13.
XX
KW sea-squirt; zinc finger protein; gene detection; drug development;
KW zinc finger protein-associated disease; gene; ds.
XX
OS Clona intestinalis.
XX
PN JP2004057126-A.
PD
PD 26-FEB-2004.
PF 31-JUL-2002; 2002JP-00222484.
XX
XX 31-JUL-2002; 2002JP-00222484.
PR
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR WPI; 2004-208711/20.
DR P-PSDB; ADP22452.
XX
XX Novel gene encoding zinc finger protein, useful as probe in gene
PT detecting instruments and in development of drug for treating zinc finger
PT protein associated diseases.
XX
PS Claim 3; SEQ ID NO 25; 972pp; Japanese.
XX
XX The invention comprises the amino acid and coding sequences of sea-squirt
CC (Clona intestinalis) zinc finger proteins. The DNA and protein sequences
CC of the invention are useful in a gene detecting instrument. The DNA and
CC protein sequences of the invention are useful in the development of drugs
CC for the treatment of zinc finger protein-associated diseases. The present
CC DNA sequence encodes a sea-squirt zinc finger protein of the invention.
XX
SQ Sequence 1372 BP; 500 A; 224 C; 286 G; 362 T; 0 U; 0 Other;
Query Match 17.2%; Score 190; DB 12; Length 1372;

Best Local Similarity 68.6%; Pred. No. 5.1e-28;		Matches 262; Conservative 0; Mismatches 120; Indels 0; Gaps 0;	
QY	7	AAAGTCGGATTTCTGTAGCCGCCAAGCCATGCGCAATAGAAATTAAGTCCAAAGGGCTCCAG	66
	55	AAACACAGGATTTCTCACTGCAAGGCTATCGGCAACCGTATCAATCAAAAGGGACTTCAA	114
QY	67	AGCCTTCGCTGCTACTGCGCAGATGTGCCAAAGCAATGCCGACGAGAAATGCTTTAAG	126
	115	AAAGTTCGGATGGTATTGTCAATGTGCCAAACACAGTGCAGAGATGAGAACGGGTTTAA	174
QY	127	TGTCACGTATGCTGAATCTCATCAAGACAACTGTGTGCTGCTTCAGAAAACCCCTCAG	186
	175	TGCCATTGATGTCGCAATACATACAGCCCAATGTTAATGCTGGTGAATCTCGGA	234
QY	187	CAGTTTATGGAATATTTTTCAGAGGAATCCGAAATGACTTTCTGGAATCTCTGAGCGGA	246
	235	GCCTTTTCACAGATTTCTTTTCAAGTTCCTTTTCAAGACTTTATGCAACTTTTGAATAACA	294
QY	247	CGCTTTGGCACTAAAGGGTCCACAACATTTGTCTACAAATGAATACATCAGCCACCGA	306
	295	CGTTTCGGGACTAAAGAGTTCAATCAATATTTGTGTATAACGAGTATATTTCCCATATA	354
QY	307	GAGCACAATCCATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAAGTGGCTG	366
	355	GAACACATTCATATGAACGCTACAAATGGGTGACGTTAACTGCTTCACAAATGGTTG	414
QY	367	GGCAGAGAGGGCTTGTGTAAG	388
	415	GGTAGAGAGGGACACTGTAAAG	436
RESULT 10			
AAC47248			
ID	AAC47248 standard; DNA; 1459 BP.		
XX	AAC47248;		
AC	AAC47248;		
XX	18-OCT-2000 (first entry)		
DT	Arabidopsis thaliana DNA fragment SRQ ID NO: 53113.		
DE	Arabidopsis thaliana		
XX	Hybridisation assay; genetic mapping; gene expression control;		
KW	protein identification; signal transduction pathway; metabolic pathway;		
KW	promoter; termination sequence; ss.		
XX	Arabidopsis thaliana		
OS	Arabidopsis thaliana		
XX	EP1033405-A2.		
PN	06-SEP-2000.		
PD	25-FEB-2000; 2000EP-00301439.		
PF	25-FEB-1999; 99US-0121825P.		
XX	05-MAR-1999; 99US-0123180P.		
PR	09-MAR-1999; 99US-0123548P.		
PR	23-MAR-1999; 99US-0125788P.		
PR	25-MAR-1999; 99US-0126264P.		
PR	29-MAR-1999; 99US-0126785P.		
PR	01-APR-1999; 99US-0127462P.		
PR	06-APR-1999; 99US-0128234P.		
PR	08-APR-1999; 99US-0128714P.		
PR	16-APR-1999; 99US-0129845P.		
PR	19-APR-1999; 99US-0130077P.		
PR	21-APR-1999; 99US-0130449P.		
PR	23-APR-1999; 99US-0130510P.		
PR	23-APR-1999; 99US-0130891P.		
PR	28-APR-1999; 99US-0131449P.		
PR	30-APR-1999; 99US-0132048P.		
PR	04-MAY-1999; 99US-0132407P.		
PR	04-MAY-1999; 99US-0132484P.		
PR	05-MAY-1999; 99US-0132485P.		

PR	06-MAY-1999;	99US-0132486P.
PR	06-MAY-1999;	99US-0132487P.
PR	07-MAY-1999;	99US-0132863P.
PR	11-MAY-1999;	99US-0134256P.
PR	14-MAY-1999;	99US-0134218P.
PR	14-MAY-1999;	99US-0134219P.
PR	14-MAY-1999;	99US-0134221P.
PR	14-MAY-1999;	99US-0134370P.
PR	18-MAY-1999;	99US-0134768P.
PR	19-MAY-1999;	99US-0134941P.
PR	20-MAY-1999;	99US-0135124P.
PR	21-MAY-1999;	99US-0135353P.
PR	24-MAY-1999;	99US-0135629P.
PR	25-MAY-1999;	99US-0136021P.
PR	27-MAY-1999;	99US-0136392P.
PR	28-MAY-1999;	99US-0136782P.
PR	01-JUN-1999;	99US-0137222P.
PR	03-JUN-1999;	99US-0137528P.
PR	04-JUN-1999;	99US-0137502P.
PR	07-JUN-1999;	99US-0137724P.
PR	08-JUN-1999;	99US-0138094P.
PR	10-JUN-1999;	99US-0138540P.
PR	10-JUN-1999;	99US-0138847P.
PR	14-JUN-1999;	99US-0139119P.
PR	16-JUN-1999;	99US-0139452P.
PR	17-JUN-1999;	99US-0139453P.
PR	17-JUN-1999;	99US-0139454P.
PR	18-JUN-1999;	99US-0139455P.
PR	18-JUN-1999;	99US-0139456P.
PR	18-JUN-1999;	99US-0139457P.
PR	18-JUN-1999;	99US-0139458P.
PR	18-JUN-1999;	99US-0139459P.
PR	18-JUN-1999;	99US-0139461P.
PR	18-JUN-1999;	99US-0139462P.
PR	18-JUN-1999;	99US-0139463P.
PR	18-JUN-1999;	99US-0139750P.
PR	18-JUN-1999;	99US-0139763P.
PR	21-JUN-1999;	99US-0139817P.
PR	23-JUN-1999;	99US-0139899P.
PR	23-JUN-1999;	99US-0140353P.
PR	24-JUN-1999;	99US-0140354P.
PR	24-JUN-1999;	99US-0140695P.
PR	28-JUN-1999;	99US-0140823P.
PR	29-JUN-1999;	99US-0140921P.
PR	30-JUN-1999;	99US-0141287P.
PR	01-JUL-1999;	99US-0141842P.
PR	02-JUL-1999;	99US-0142055P.
PR	06-JUL-1999;	99US-0142390P.
PR	08-JUL-1999;	99US-0142803P.
PR	09-JUL-1999;	99US-0142920P.
PR	12-JUL-1999;	99US-0142977P.
PR	13-JUL-1999;	99US-0143542P.
PR	14-JUL-1999;	99US-0143624P.
PR	15-JUL-1999;	99US-0144005P.
PR	16-JUL-1999;	99US-0144085P.
PR	19-JUL-1999;	99US-0144086P.
PR	19-JUL-1999;	99US-0144325P.
PR	19-JUL-1999;	99US-0144331P.
PR	19-JUL-1999;	99US-0144332P.
PR	19-JUL-1999;	99US-0144333P.
PR	19-JUL-1999;	99US-0144334P.
PR	19-JUL-1999;	99US-0144335P.
PR	20-JUL-1999;	99US-0144352P.
PR	20-JUL-1999;	99US-0144632P.
PR	20-JUL-1999;	99US-0144884P.
PR	21-JUL-1999;	99US-0144814P.
PR	21-JUL-1999;	99US-0145086P.
PR	21-JUL-1999;	99US-0145088P.
PR	22-JUL-1999;	99US-0145085P.
PR	22-JUL-1999;	99US-0145087P.

XX 28-APR-2003; 2003US-00425114.
PF
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
PA (LIUJ/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOV/) CAO Y.
XX
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX
XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
PS Claim 1; SEQ ID NO 2975; 15pp; English.
XX
CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
SQ Sequence 1614 BP; 399 A; 362 C; 493 G; 360 T; 0 U; 0 Other;
Query Match 15.3%; Score 168.8; DB 13; Length 1614;
Best Local Similarity 64.7%; Pred. No. 8.2e-24;
Matches 251; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 1 ATGGGCAAGTCGGATTCTTGAGCCCAAGCCCATGCCAATAGATTAAGTCCAAAGGG 60
DB 80 ATGGGGAAGCAGCAGTTCCTGACCCGAGCGCATCGCAAAACCGGATCAAGGCGAAGGG 139
QY 61 CTCAGAGCTTCGCTGTAAGTCTGTAATCTCATCAAGACAACCTGTTGCGTTCAGAAAC 120
DB 140 CTGAGAGCTTCGCGTGATGTCGATGTCGAGAGCAGTCCCGCAGAGAACGG 199
QY 121 TTTAAGTGTCATGATGTCGTAATCTCATCAAGACAACCTGTTGCGTTCAGAAAC 180
DB 200 TTCAAGTGCCATGTCGAGTTCGACCCAGAGGCGAGTGCAGGTTTCGCGCATGCC 259
QY 181 CCTCAGAGTTTATGGAATATTTTTCAGAGAAATTCGAAATGATCTTCTGGAACCTCTG 240
DB 260 CCCGACCGGTCGTCGAGGGTCTCCGAGGAGTTCCTCGAGTCTCTTCTCTCCCTCATC 319
QY 241 AGGCGAGCTTGGCACTAAAGGTCACAAACACATGCTCAATGAATATCATCAGC 300
DB 320 CGCCGCGCGCACCGCCACTCCCGCGTCGCCGCCACCGTCTGTCACAGGATCATGCC 379
QY 301 CACCGAGAGCATCCACATGAAGCGTACCAGTGGGAGACACTGACCGACTTTACCAAG 360
DB 380 GACCGGACCACTCCACATGAATCCACGGGTGGGCGGACGCTCACCGAGTTCGTCAAG 439

QY 361 TGGCTGGGCAGAGAGGGCTTGTGTAAG 388
DB 440 TTCCTGGGCGCGAGGGGTACTGTAAGG 467
RESULT 12
AD47874
ID AD47874 standard; cDNA; 1713 BP.
XX
AC AD47874;
XX
DT 21-APR-2005 (first entry)
XX
XX Plant full length insert polynucleotide seqid 22614.
XX
XX plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content; gene; ss.
XX
OS Unidentified.
XX
PN US2004034888-A1.
XX
PD 19-FEB-2004.
XX
PF 28-APR-2003; 2003US-00425114.
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
XX (LIUJ/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOV/) CAO Y.
XX
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX
XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
PS Claim 1; SEQ ID NO 22614; 15pp; English.
XX
CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
SQ Sequence 1713 BP; 423 A; 422 C; 504 G; 364 T; 0 U; 0 Other;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:58:53 ; Search time 6160.01 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-2
Perfect score: 1102
Sequence: 1 agggcagctcgatttctt.....aaaaaaaaaaaaaaaaaaaa 1102

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_gss1.*
10: gb_gss2.*
11: gb_gss3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	625.6	56.8	677	2	BF321724
C 2	576	52.3	1081	2	BE283156
C 3	559	50.7	560	2	BG072126
C 4	554.2	50.3	739	5	BQ782294
C 5	545.6	49.5	754	5	BQ780218
C 6	522	47.4	601	3	BP774644
C 7	512	46.5	512	3	BM228037
C 8	512	46.5	512	3	BM229452
C 9	496.2	45.0	1154	10	AY412520
C 10	474	43.0	1396	4	CS595908
C 11	462	41.9	513	2	BE225867
C 12	454	41.2	466	2	BS448684
C 13	449	40.7	1543	4	AY609991
C 14	448.2	40.7	457	1	AW045702
C 15	433.4	39.3	447	2	BB749127
C 16	433.4	39.3	864	1	AM037867
C 17	431	39.1	723	3	BQ186497
C 18	429.8	39.0	871	1	AJ819723
C 19	426.6	38.7	526	2	BS114414
C 20	424.4	38.5	430	6	CF617307
C 21	423.8	38.5	800	1	AM038283
C 22	414	37.6	437	2	BB749400

23	413	37.5	682	1	AV721396
24	406.4	36.9	679	1	AM028238
C 25	404	36.7	733	5	BX104466
26	402.8	36.6	786	1	AJ819816
27	401	36.4	703	1	AM035210
C 28	400	36.3	493	2	BE120977
C 29	398.4	36.2	691	1	AM028686
C 30	397.4	36.1	848	1	AL558810
C 31	397	36.0	754	8	DN755964
32	396	35.9	543	6	CB158644
33	395.4	35.9	399	5	BY395138
C 34	392.6	35.6	720	7	CK301009
C 35	389.8	35.4	820	6	CF409895
36	388	35.2	638	4	AK009429
37	386.4	35.1	625	7	CF914851
38	384	34.8	501	6	CA560899
39	384	34.8	538	6	CA559106
40	383.4	34.8	515	1	AA638405
41	383.2	34.8	583	2	BF321725
42	383.2	34.8	663	1	AJ660240
C 43	383	34.8	698	1	AI650375
C 44	382.4	34.7	1077	5	BUS11663
C 45	379	34.4	391	1	AU041571

ALIGNMENTS

RESULT 1
BF321724/c
LOCUS
DEFINITION
uz64C04.x1 NCI CGAP Mam6 Mus musculus cDNA clone IMAGE:3673830 3', similar to TR:060870 O60870 KIN17 PROTEIN. ; mRNA sequence.
ACCESSION
BF321724
VERSION
BF321724.1
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
1 (bases 1 to 677)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml
MGI:1434598
High quality sequence stop: 396.
FEATURES
Location/Qualifiers
1..677
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3673830"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator

ORIGIN		providing samples: Jeffrey Green, M.D., NIH		COMMENT	
Query Match		56.8%;	Score 625.6;	DB 2;	Length 677;
Best Local Similarity		96.3%;	Pred. No. 3.1e-124;		
Matches		653;	Conservative	0;	Mismatches 19; Indels 6; Gaps 1;
QY	388	GGTACAGTGCACTGAAGCTGTCTGGGAGCGCAGCATCCGGGAAACGGAAGAGTCTTCA	447		
Db	677	GGCCCAAGTGCACTGAAGCTGTCTGGGAGCGCAGCATCCGGGAAACGGAAGAGTCTTCA	618		
QY	448	CAGAGCTCGGCCAGCTCGGAGAGAGAGTCCGCCCTGGATGAGATCATGAGCTC	507		
Db	617	CAGAGCTCGGCCAGCTCGGAGAGAGAGTCCGCCCTGGATGAGATCATGAGCTC	558		
QY	508	GAAGAGGAAAGAAAGGACCGCAGCGACAGACGCTCGTTACAGCCGGGATCGTTGTG	567		
Db	557	GAAGAGGAAAGAAAGGACCGCAGCGACAGACGCTCGTTACAGCCGGGATCGTTGTG	498		
QY	568	AAATATTAAACGAAGAAGCTTGGGAGAAATATCAAGAAGAAAGGGTCTGTTAAGGAA	627		
Db	497	ACAATTTAAACGAAGAAGCTTGGGAGAAATATCAAGAAGAAAGGGTCTGTTAAGGAA	438		
QY	628	GTGATTGACAGGTACACAGCTGTGTTAAGATGACTCTCTGGACAGGCTGAAACTG	687		
Db	437	GTGATTGACAGGTACACAGCTGTGTTAAGATGACTCTCTGGACAGGCTGAAACTG	378		
QY	688	GACGAGACTCATTTAGAGACAGTCATTCGGGCCCGGGGAAAGGGTCTAGTTTTAAAT	747		
Db	377	GACGAGACTCATTTAGAGACAGTCATTCGGGCCCGGGGAAAGGGTCTAGTTTTAAAT	318		
QY	748	GGAGGCTACAGAGGAAATGAAGCACTCTCGAATCCATCAATGAGAGGCTTTTTCAGCC	807		
Db	317	GGAGGCTACAGAGGAAATGAAGCACTCTCGAATCCATCAATGAGAGGCTTTTTCAGCC	258		
QY	808	ACGATAGTCATTTGAAGCTGCTTGAAGAGCGCAGAGTTGAAGGTATTCATATGAA	867		
Db	257	ACGATAGTCATTTGAAGCTGCTTGAAGAGCGCAGAGTTGAAGGTATTCATATGAA	198		
QY	868	GACATATCTAAACTTCCTTGAGTTTGAAATTTGATAACACACATTTGAACTGTGAAGC	927		
Db	197	GACATATCTAAACTTCCTTGAGTTTGAAATTTGATAACACACATTTGAACTGTGAAGC	138		
QY	928	ATCAAAATGGTGTAGCCAGGCACTGTGTAACCTACTGTGTAGGGGATTTGTTTGT	987		
Db	137	ATCAAAATGGTGTAGCCAGGCACTGTGTAACCTACTGTGTAGGGGATTTGTTTGT	78		
QY	988	ATTAAAAAATAAATCATCTATTAAATACCTAGTGAATAGTTGGGTAATTTAATA	1047		
Db	77	AT-----AAAAAATAATCATCTATTAAATACCTAGTGAATAGTTGGGTAATTTAATA	24		
QY	1048	AAATCTATGTTTTTTTTTA 1065			
Db	23	AAATCTATGTTTTTTAA 6			
RESULT 2					
BE283156					
LOCUS		601103709F1 NCI_CGAP_Lu29 Mus musculus cDNA clone IMAGE:3500835 5',			
DEFINITION		mRNA sequence.			
ACCESSION		BE283156			
VERSION		BE283156.1 GI:9159216			
KEYWORDS		EST.			
SOURCE		Mus musculus (house mouse)			
ORGANISM		Mus musculus			
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.			
REFERENCE		1 (bases 1 to 1081)			
AUTHORS		NIH-MGC http://mgc.nci.nih.gov/			
TITLE		National Institutes of Health, Mammalian Gene Collection (MGC)			
JOURNAL		Unpublished (1999)			

Contact: Robert Strausberg, Ph.D.
Email: cgabs@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM8560 row: b column: 04
High quality sequence stop: 641.

Location/Qualifiers

1..1081

/organism="Mus musculus"

/mol_type="mRNA"

/strain="Czech II"

/db_xref="taxon:10090"

/clone="IMAGE:3500835"

/tissue_type="spontaneous tumor, metastatic to mammary."

Stem cell origin."

/lab_host="DH10B"

/clone_lib="NCI CGAP Lu29"

/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"

ORIGIN

Query Match

Best Local Similarity

Matches

639;

Conservative

0;

Mismatches

65;

Indels

4;

Gaps

2;

QY

388

GGTACAGTGCACTGAAGCTGTCTGGGAGCGCAGCATCCGGGAAACGGAAGAGTCTTCA

447

Db

63

GGACCAAGTGCACTGAAGCTGTCTGGGAGTGCGAGCATCCGGGAAACGGAAGAGTCTTCA

122

QY

448

CAGAGCTCCGCCAGCTCGGAGAGAGAGTCCGCCCTGGATGAGATCATGAGCTC

507

Db

123

CAGAGCTCCGCCAGCTCGGAGAGAGAGTCCGCCCTGGATGAGATCATGAGCTC

182

QY

508

GAAGAGGAAAGAAAGGACCGCAGCGACAGACGCTCGTTACAGCCGGGATCGTTGTG

567

Db

183

GAAGAGGAAAGAAAGGACCGCAGCGACAGACGCTCGTTACAGCCGGGATCGTTGTG

242

QY

568

AAATTTATAACGAAGAAGCTTGGGAGAAATATCAAGAAGAAAGGGTCTGTTAAGGAA

627

Db

243

AAATTTATAACGAAGAAGCTTGGGAGAAATATCAAGAAGAAAGGGTCTGTTAAGGAA

302

QY

628

GTGATTGACAGGTACACAGCTGTGTTAAGATGACTGACTCTGGAGACAGGCTGAAACTG

687

Db

303

GTGATTGACAGGTACACAGCTGTGTTAAGATGACTGACTCTGGAGACAGGCTGAAACTG

362

QY

688

GACGAGCTCATTTAGAGACAGTCATTCGGGCCCGGGGAAAGGGTCTAGTTTTAAAT

747

Db

363

GACGAGCTCATTTAGAGACAGTCATTCGGGCCCGGGGAAAGGGTCTAGTTTTAAAT

422

QY

748

GGAGGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCAATGAGAGGCTTTTTCAGCC

807

Db

423

GGAGGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCAATGAGAGGCTTTTTCAGCC

482

QY

808

ACGATAGTCATTTGAAGCTGAGACCTTTGAAAGGACGCGAGAGTTCAGAGTATCAATATGAA

867

Db

483

ACGATAGTCATTTGAAGCTGAGACCTTTGAAAGGACGCGAGAGTTCAGAGTATCAATATGAA

542

QY

868

GACATATCTAAACTTCCTTGAGTTTGAAATTTGATAACAA--CACATTTGAAACTGTGAA

925

Db

543

GACATATCTAAACTTCCTTGAGTTTGAAATTTGATAACAAACACATTTGAAACTGTGAA

602

QY

926

GCATCAAAATGGTGTAGCCAGGACCTGTGTAACTCTACTGTGTAGGGGATTTGTTTT

985

Db

603

GCATCAAAATCGATGTTAGCCAGGCACTGTGTAACTCTACTGTGGGTGAGGACTTCGGTC

662

QY

986

GTATTTAAAAAATAAATAATCATCTATTAAATACCTAGTGAATAGTTGGGTAAATTTAATA

1045

```

663 TGACTTAACAAATAGGTCATCTATTACATCTTGTGGAGTAGTGGCTAACTTATTA- 721
1046 TAAAACTATGTTTTTTTAAAGTGTAATAAAAAAAGAAAAA 1093
722 -TAACTATGTTGTTTAAACCGAAACCAACAAATATAGGAAAAA 768

RESULT 3
BG072126/c
LOCUS
DEFINITION
H3107A09-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
H3107A09 3', mRNA sequence.
ACCESSION
BG072126 GI:40072826
VERSION
BG072126.2
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.
1 (bases 1 to 560)
Tanaka, T.S., Jaradat, S.A., Lim, M.K., Kargul, G.J., Wang, X.,
Grabovac, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H.,
Wood, W.H. III, Becker, K.G. and Ko, M.S.H.
Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
10922068
On Jan 26, 2001 this sequence version replaced gi:12554695.
Other ESTs: H3107A09-5
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://igun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3107 row: A column: 09
Seq primer: -21M13 Forward
High quality sequence stop: 560
POLYA=Yes,
Location/Qualifiers
1..560
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="nlaEST:H3107A09-3"
/db_xref="taxon:10090"
/clones="H3107A09"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/note="vector: pSFR1; Site1: SalI, Site2: NotI; This
clone is among a rearranged set of 15,247 clones from 11
embryo cDNA libraries (including preimplantation stage
embryos from unfertilized egg to blastocyst, embryonic
part of E7.5 embryos, extraembryonic part of E7.5
embryos, and E12.5 female mesonephros/gonad) and one
newborn ovary cDNA library. Average insert size 1.5 kb.
All source libraries are cloned unidirectionally with
Oligo (dT)-Not primers. References include: (1)
Genome-wide expression profiling of mid-gestation
placenta and embryo using a 15,000 mouse developmental
cDNA microarray, 2000, Proc. Natl. Acad. Sci. U.S.A. 97:
9127-9132; (2) Large-scale cDNA analysis reveals phased
gene expression patterns during preimplantation mouse
development, 2000, Development, 127: 1737-1749; (3)
Genome-wide mapping of unselected transcripts from
extraembryonic tissue of 7.5-day mouse embryos reveals
enrichment in the t-complex and under-representation on
the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

Query Match 50.7%; Score 559; DB 2; Length 560;
Best Local Similarity 100.0%; Pred. No. 7.1e-110; Indels 0; Gaps 0;
Matches 559; Conservative 0; Mismatches 0;

QY 512 AGGAAAGAAAAAGGACCGCACGACAGACGCTGTTACAGCCGGGGATCGTTGTGAAAA 571
DB 560 AGGAAAGAAAAAGGACCGCACGACAGACGCTGTTACAGCCGGGGATCGTTGTGAAAA 501
QY 572 TTATAACGAAGAAGCTTGGGGAGAAATATCAACAAGAAAGGGTCTGTTAGGAAGTGA 631
DB 500 TTATAACGAAGAAGCTTGGGGAGAAATATCAACAAGAAAGGGTCTGTTAGGAAGTGA 441
QY 632 TTGACAGGTACACAGCTGTGTTAAAGATGACTCTCGAGACAGCGCTGAAACTCGACC 691
DB 440 TTGACAGGTACACAGCTGTGTTAAAGATGACTCTCGAGACAGCGCTGAAACTCGACC 381
QY 692 AGACTCATTTAGAGACAGTCATTCCGGCCCCCGGGAAAAAGGGTTCCTAGTTTAAATGGAG 751
DB 380 AGACTCATTTAGAGACAGTCATTCCGGCCCCCGGGAAAAAGGGTTCCTAGTTTAAATGGAG 321
QY 752 GTTACAGAGAAATGAAGGCACCTCTCGAATCCATCAATGAGAAGGCTTTTTCAGCCACGA 811
DB 320 GTTACAGAGAAATGAAGGCACCTCTCGAATCCATCAATGAGAAGGCTTTTTCAGCCACGA 261
QY 812 TAGTCATTGAAACTGCGACCTTTGAAAGGACGACGAGTTCGAAGGTATTCATATGAAGACA 871
DB 260 TAGTCATTGAAACTGCGACCTTTGAAAGGACGACGAGTTCGAAGGTATTCATATGAAGACA 201
QY 872 TATCTAAACTTGCTTGAGTTTGAAAAATTTGATAACACACATTTGAAACTGTGGAAGCATCA 931
DB 200 TATCTAAACTTGCTTGAGTTTGAAAAATTTGATAACACACATTTGAAACTGTGGAAGCATCA 141
QY 932 AATTGCTGTAGCAAGGACCTGTGTAACCTCTACTCTGTGTAGGGGATTTGTTTGTATTA 991
DB 140 AATTGCTGTAGCAAGGACCTGTGTAACCTCTACTCTGTGTAGGGGATTTGTTTGTATTA 81
QY 992 AAAAAAANAATCATCTATTTAAATCTAGTGAATAGTTGGTAAATTTATATAAANAAT 1051
DB 80 AAAAAAANAATCATCTATTTAAATCTAGTGAATAGTTGGTAAATTTATATAAANAAT 21
QY 1052 CTATGTTTTTTTTTAAGTGT 1070
DB 20 CTATGTTTTTTTTTAAGTGT 2

RESULT 4
BG072126/c
LOCUS
DEFINITION
H3107A09-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
H3107A09 3', mRNA sequence.
ACCESSION
BG072126
VERSION
BG072126.2
KEYWORDS
EST.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
REFERENCE
1 (bases 1 to 739)
Bonalido, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

```

Email: bento-soares@uiowa.edu
 Tissue Procurement: Jeff Stevens
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: DISTRIBUTION: Researchers may obtain clones
 from Research Genetics (www.resgen.com).
 The following repetitive elements were found in this cDNA
 sequence: 1-47, >AT-rich#Low_complexity (matched complement)
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

Location/Qualifiers
 1..739
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="UI-R-FF0-cpj-b-08-0-UI"
 /tissue_type="Mixed tissues"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Vector: pT7T3-Pac (Pharmacia) with a modified
 polylinker; Site 1: EcoR I; Site 2: Not I; UI-R-FF0 is a
 subtracted cDNA library containing the following
 tissue(s): Normal cartilage and SR-JWS Tumor Line . The
 subtraction was made according to Bonaldo, Lennon and
 Soares, Genome Research, 6:791-806, 1996. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for these libraries are: CTAATGGAGC,
 CATTCTTGTA.
 TAG_TISSUE=rat SRC-JWST tumor line
 TAG_L1B=UI-R-FF0
 TAG_SEQ=CATTCTTGTA"

ORIGIN

Query Match 50.3%; Score 554.2; DB 5; Length 739;
 Best Local Similarity 90.2%; Pred. No. 7.9e-109;
 Matches 632; Conservative 0; Mismatches 53; Indels 16; Gaps 3;

QY 388 GTTACAGTGCCTGAAGCTGCTGGGAGCGAGCATCCGGAAACGGAAGATCTTCA 447
 DB |||||
 QY 692 GAACGAGCGCACTGAAGATGCTGGGAGCAGCATCCGTGAGCGGAAGATCTTCA 633
 DB |||||
 QY 448 CAGAGTCCGCCAGCCTGC---GAAGAAGAAGAGTCGCGCCTGGATGAGATCATGGAG 504
 DB |||||
 QY 632 CAGAGTCAAGTCAGCTCGAAGAAGAAGAGAGTCTGCCCTGGATGAGATCATGGAG 573
 QY 505 CTCGAAGAGAAAGAAAGACCGCAGCAGCATCGCTGGTTACGCGGGGATCGTT 564
 DB |||||
 QY 572 ATTGAGGAGAAAGAAAGAGACCGCAGCAGCATCTCTGGTTACAGCTCGGATCAT 513
 QY 565 GTGAAATTTATACGAAGAGCTTGGGAGAGATATCACAAGAAGAGGGTCTGTAG 624
 DB |||||
 QY 512 GTGAAATTTATACGAAGAGCTTGGGAGAGATATCACAAGAAGAGGGTCTGTAG 453
 QY 625 GAAGTGATTGACAGGTACACAGCTGTGGTAAAGATGACTCTCTGGAGACAGGCTGAA 684
 DB |||||
 QY 452 GAAGTGATTGACAGGTACACAGCTGTGGTAAAGATGACTCTCTGGAGACAGGCTGAA 393
 QY 685 CTGGACAGACTCATTTAGAGACAGTCATTCCGGCCCGGGGAAAGAGGTTCTAGTTTA 744
 DB |||||
 QY 392 CTGGACAGACTCATTTAGAGACAGTCATTCCAGCCCGGAAAGAGGTTCTAGTTTA 333
 QY 745 AATGAGGCTACAGAGGAATGAAGCACTCTCGAATCCATCAATGAGAGGCTTTTCA 804
 DB |||||
 QY 332 AACGAGGCTACAGAGGAATGAAGCACTCTCGAATCCATCAATGAGAGGCTTTTCA 273
 QY 805 GCCAGATAGTCATTGAAACTGGACCTTTTGAAGGACGAGAGGTTGAAGGATTTCAATAT 864
 DB |||||
 QY 272 GCCAGATAGTCATTGAAACGGGACCTTTTGAAGGCGCAGAGTTGAAGGATTTCAATAT 213

QY 865 GNAGACATATCAACTTGGCTTGGAGTTTGAAATTTGATAACACACACATTGAAACTGTCA 924
 DB |||||
 QY 212 GAAGACATATCTAAACTTGGCTTGGAGTTTGAAATTTGATAACACACATTGAAATGTAA 153
 QY 925 AGCATCAAAATTTGGTTAGCCAAAGCAGCTGTGTAACTCTACTGTGTAGGGGATTTGTT 984
 DB |||||
 QY 152 AGCATCAAAATTTGAAGTTAGCCAGCAGCTGTGTAACTCTACTGTGTAGGGGATTTGTT 93
 QY 985 TGTATTAAAAAATAAATCATCTATTAAATAGTGTGAATAG-----TTGGGTAAAT 1040
 DB |||||
 QY 92 -----ATAAGAAAGTCACTATTATAATACCACTGGATAGCTGTTGGGTAAAT 42
 QY 1041 TATAATAAATCTATGTTTTTTTAAAGTGTAAAAAATAA 1081
 DB |||||
 QY 41 TATAATAAATCTAATTTTTTTTTTAAAAAATAAATAA 1

RESULT 5
 BQ780218/c
 LOCUS
 DEFINITION
 BQ780218
 ACCESSION
 BQ780218
 VERSION
 BQ780218.1 GI:21988690
 KEYWORDS
 EST.
 SOURCE
 Rattus norvegicus (Norway rat)
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridea; Muridae; Murinae; Rattus.
 1 (bases 1 to 754)
 REFERENCE
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL
 Genome Res. 6 (9), 791-806 (1996)
 PUBMED
 8889548
 COMMENT
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Jeff Stevens
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: DISTRIBUTION: Researchers may obtain clones
 from Research Genetics (www.resgen.com).
 The following repetitive elements were found in this cDNA
 sequence: 1-54, >AT-rich#Low_complexity (matched complement)
 Seq primer: M13 FORWARD
 POLYA=Yes.

Location/Qualifiers
 1..754
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /db_xref="taxon:10116"
 /clone="UI-R-FF0-cpb-m-06-0-UI"
 /tissue_type="Mixed tissues"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Vector: pT7T3-Pac (Pharmacia) with a modified
 polylinker; Site 1: EcoR I; Site 2: Not I; UI-R-FF0 is a
 subtracted cDNA library containing the following
 tissue(s): Normal cartilage and SR-JWS Tumor Line . The
 subtraction was made according to Bonaldo, Lennon and
 Soares, Genome Research, 6:791-806, 1996. The
 oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)18 tail. The
 sequence tags for these libraries are: CTAATGGAGC,

FEATURES
 source


```

CATCTTGTA.
TAG_TISSUE=cartilage
TAG_LIB=UI-R-PFO
TAG_SEQ=CTAATGGAGC"

ORIGIN
  Query Match      49.5%; Score 545.6; DB 5; Length 754;
  Best Local Similarity 39.7%; Pred. No. 5.7e-107;
  Matches 635; Conservative 0; Mismatches 56; Indels 17; Gaps 4;

  388 GGTACACAGTGCACAGCTGCTGGGAGCGCAGCATCCGGGAAACGGAAGAGTCTTCA 447
  698 GGACCGAGCGCACTGAAGATGCT-GGAGCACAGCATCCGTGAGCGGAAAGAGTCTTCA 640
  448 CAGAGCTCCGCCAGCCTGC---GAAGAAGAAGTCGCGCCCTGGATGATGATCATGGAG 504
  639 CAGAGCTCAGCTCAGCTCGGAGGAAGAAGAAGTCTGCGCTGGATGATCATGGAG 580
  505 CTCGAGAGGAAAGAAAGAGCCGACGAGCAGAGCCCTGGTTACAGCCGGGATCGTT 564
  579 ATTGAGGAGGAAGAAGAAGAGCCGACGAGCAGACTCTCTGTTACAGCTGGGATCAT 520
  565 GTGAAAATTATAACGAAGAAGCTTTGGGAGAGAAATATCACAAAGAAGAAAGGGTCTTAAG 624
  519 GTGAAAATTATAACGAAGAAGCTTTGGGAGAGAAATACCACAAAGAAGAAAGGGTCTTAAG 460
  625 GAAGTGATGACAGGTAACAGCTGTGTTAAGATGACTGACTCTCGGAGACAGCTGAA 684
  459 GAAGTGATGACAGATACACAGCTGTGTTAAGATGACTGACTCTCGGAGACAGCTGAA 400
  685 CTGACACAGACTCATTTAGAGACAGTCAATCCGCGCCCGGGGAAAGGGTCTTAGTTTA 744
  399 CTGACACAGACTCATTTAGAGACAGTCAATCCAGCCCCAGGAAAGAGTCTAGTTTA 340
  745 AATGGAGGCTACAGAGAAATGAAGCAGCTCTGAAATCCATCAATGAAGAGGCTTTTCA 804
  339 AACGGAGCTACAGAGAAATGAAGCAGCCCTCGAGTCCATCAATGAAGAGCTTTTCA 280
  805 GCCACGATAGTCAATGAAACCTGTTGAAGGAGCGCAGAGTTGAAGTATTCATAT 864
  279 GCCACGATAGTCAATGAAACCTGTTGAAGGAGCGCAGACTTGAAGGTATTCAGTAT 220
  865 GAAGACATATCAACTGCTTGAGTTGAAATTTGATAACACACATTTGAACACTGTA 924
  219 GAAGACATATCAACTGCTTGAGTTGAAATTTGATAACACATTTGATAACACTGTA 160
  925 AGCATCAAAATGGTGTAGCAAGCAGCTGTGTAACCTCTACTGTGTTAGGGGATTTGTT 984
  159 AGCATCAAAATGGTGTAGCAAGCAGCTGTGTAACCTCTACTGTGTTAGGGGATTTGTT 100
  985 TGTATTAATAAATAAATAAATCATCTATTAAATACTAGTGAATAG-----TTGGGTAATTT 1040
  99 -----ATAAGAAGATNATCTATTAAATACCACCTGGATAGCTGTTTGGGTAAAT 49
  1041 TATAATAAATCTATGTTTTTTTAAAGTGTAATAAATAAATAAATAAATAAATAA 1088
  48 TATAATAAATCTAATTTTAAAGTGTCATAAATAAATAAATAAATAAATAAATAA 1

RESULT 6
BP774644/c
LOCUS
DEFINITION
  BP774644 mouse (C57BL/6) pancreatic islet library with
  recombination-based method Mus musculus cDNA clone mif14029 3',
  mRNA sequence.
ACCESSION
  BP774644
VERSION
  BP774644.1 GI:50233342
KEYWORDS
  EST.
SOURCE
  Mus musculus (house mouse)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muroidea; Muridae; Murinae; Mus.

```

```

1 (bases 1 to 601)
Nishimura,M., Yokoi,N., Miki,T., Horikawa,Y., Yoshioka,H.,
Takeda,J., Ohara,O. and Seino,S.
Construction of a multi-functional cDNA library specific for mouse
pancreatic islets and its application to microarray
DNA Res. 11 (5), 315-323 (2004)
15747579
Contact: Susumu Seino
Division of Cellular and Molecular Medicine
Kobe University Graduate School of Medicine
7-5-1 Kusunoki-cho, Chuo-ku, Kobe, Hyogo 650-0017, Japan
Tel: 81-78-382-5360
Fax: 81-78-382-5370
Email: seino@med.kobe-u.ac.jp.
Location/Qualifiers
  1..601
  /organism="Mus musculus"
  /mol_type="mRNA"
  /strains="C57BL/6"
  /db_xref="taxon:10090"
  /clone="mif14029"
  /sex="male"
  /tissue_type="pancreatic islet"
  /dev_stage="adult"
  /clone_lib="mouse"
  recombination-based method"

ORIGIN
  Query Match      47.4%; Score 522; DB 3; Length 601;
  Best Local Similarity 95.5%; Pred. No. 6.8e-102;
  Matches 548; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

  509 AAGAGAAAGAAAGAGCCGACGACAGACGCTTGGGGAGAAATATCACAAAGAAGAAAGGGTCTTAAAGAAAG 568
  591 AAGAGAAAGAAAGAGCCGACGACGCTTGGGGAGAAATATCACAAAGAAGAAAGGGTCTTAAAGAAAG 533
  569 AAATTATAACGAAGAAGCTTTGGGGAGAAATATCACAAAGAAGAAAGGGTCTTAAAGAAAG 628
  532 AAATTATAACGAAGAAGCTTTGGGGAGAAATATCACAAAGAAGAAAGGGTCTTAAAGAAAG 473
  629 TGATTGACAGGTACACAGCTGTGTTAAGATGACTGACTCTGGAGACAGGCTGAAACTGG 688
  472 TGATTGACAGGTACACAGCTGTGTTAAGATGACTGACTCTGGAGACAGGCTGAAACTGG 413
  689 ACCAGACTCATTTAGAGACAGTCAATCCGCGCCCGGGGAAAGGGTCTTAGTTTAAATG 748
  412 ACCAGACTCATTTAGAGACAGTCAATCCGCGCCCGGGGAAAGGGTCTTAGTTTAAATG 353
  749 GAGGCTACAGAGGAAATGAAGCAGCTCTCGAATCCAATCAATGAAGAGGCTTTTTCAGCCA 808
  352 GAGGCTACAGAGGAAATGAAGCAGCTCTCGAATCCAATCAATGAAGAGGCTTTTTCAGCCA 293
  809 CGATAGTCAATGAAACTGGACCTTTGAAAGGAGCGCAGAGTTGAAGGTATTCATATGAAG 868
  292 CGATAGTCAATGAAACTGGACCTTTGAAAGGAGCGCAGAGTTGAAGGTATTCATATGAAG 233
  869 ACATATCTAAACTGCTTGAGTTGAAATTTGATAACACACATTTGAAACTCTGTAAGCA 928
  232 ACATATCTAAACTGCTTGAGTTGAAATTTGATAACACACATTTGAAACTCTGTAAGCA 173
  929 TCAAAATGGTGTAGCAAGGCACTGTGTAACCTCTACTGTGTAGGGGATTTGTTTGTGTA 988
  172 TCAAAATGGTGTAGCAAGGCACTGTGTAACCTCTACTGTGTAGGGGATTTGTTTGTGTA 113
  989 TTAATAAATAAATAAATCATCTATTAAATACTAGTGAATAGTTGGGTAAATTTTATAATAA 1048
  112 TTAATAAATAAATAAATCATCTATTAAATACTAGTGAATAGTTGGGTAAATTTTATAATAA 53
  1049 AATCTATGTTTTTTTTTAAAGTGTAATAAATAAATAAATAAATAAATAAATAA 1082
  52 AATTTATTTTTTTTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 19

```

RESULT 7
BM228037/c
LOCUS
DEFINITION
Musculus cDNA clone NIA:K0254C06 IMAGE:30049853 3', mRNA sequence.
ACCESSION
BM228037.2 GI:31485640
VERSION
EST.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
TITLE
Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)
On Dec 14, 2001 this sequence version replaced gi:17790538.
JOURNAL
PUBMED
11544199
COMMENT
Other ESTs: K0254C06-5N
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@gsun.grc.nia.nih.gov
Plate: K0254 row: C column: 06
Seq primer: -21M13 Forward
High quality sequence stop: 512
POLYA=Yes.
FEATURES
source
1. .512
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="C57BL/6J"
/db_xref="niaEST:K0254C06-3"
/cloned="NIA:K0254C06 IMAGE:30049853"
/tissue_type="Unfertilized Egg"
/lab_host="DH10B"
/clone_lib="NIA Mouse Unfertilized Egg cDNA Library
(Long)"
/notes="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
extracted from a pool of 1488 unfertilized eggs.
Double-stranded cDNAs were synthesized with an Oligo(dT)
primer [Invitrogen].
5'-pGACTAGTCTCATGCGAGCGCGCCCTTTTCTTTT-3',
treated with T4 DNA polymerase, and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.5 kb. The library was constructed
by Yulan Piao (NIA)."
ORIGIN
Query Match 46.5%; Score 512; DB 3; Length 512;
Best Local Similarity 100.0%; Pred. No. 9.7e-100; Indels 0; Gaps 0;
Matches 512; Conservative 0; Mismatches 0;
479 AGTCGGCCCTGGATGAGATCATGGAGCTCGAAGAGGAAAGAAAGACCGCAGCAG 538

```
/db xref="taxon:10090"
/clone="NIA:K0278H05 IMAGE:30052216"
/tissue_type="Unfertilized Egg"
/lab_host="DH10B"
/clone_lib="NIA Mouse Unfertilized Egg cDNA Library
(Long)"
/notes="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://igsn.gri.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
extracted from a pool of 1488 unfertilized eggs.
Double-stranded cDNAs were synthesized with an Oligo(dT)
primer [Invitrogen]:
5'-pGACTAGTCTAGATCCGAGCGGCCCTTTTTTTTTTTT-3',
treated with T4 DNA polymerase, and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker L1-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.5 kb. The library was constructed
by Yulan Piao (NIA)."
```

ORIGIN

Query Match 46.5%; Score 512; DB 3; Length 512;
Best Local Similarity 100.0%; Pred. No. 9.7e-100; Indels 0; Gaps 0;
Matches 512; Conservative 0; Mismatches 0

479 AGTCGGCCTGGATGAGATCATGAGCTCGAAGAGGAGAAAGAACCGACGACAG 538
Db |
512 AGTCGGCCTGGATGAGATCATGAGCTCGAAGAGGAGAAAGAACCGACGACAG 453
Qy |
539 ACGCCTGGTTACACCGCGGGATCGTTGTGAAAATTATAACGAAGAAGCTTGGGAGAAAT 598
Db |
452 ACGCCTGGTTACACCGCGGGATCGTTGTGAAAATTATAACGAAGAAGCTTGGGAGAAAT 393
Qy |
599 ATCAACAAGAAAGGGGTGTTAAGGAAGTGTGACAGTACACAGCTGTGTTAAGA 658
Db |
392 ATCAACAAGAAAGGGGTGTTAAGGAAGTGTGACAGTACACAGCTGTGTTAAGA 333
Qy |
659 TGACTGACTCTGGAGACAGCGCTGAACTGGACACAGACTCATTTAGACAGACTCATTCGG 718
Db |
332 TGACTGACTCTGGAGACAGCGCTGAACTGGACACAGACTCATTTAGACAGACTCATTCGG 273
Qy |
719 CCCCAGGAAAGGGTCTAGTTTTAAATGAGGCTACAGAGGAAATGAAGGCACTCTCG 778
Db |
272 CCCCAGGAAAGGGTCTAGTTTTAAATGAGGCTACAGAGGAAATGAAGGCACTCTCG 213
Qy |
779 AATCATCAATGAGAGGCTTTTTCAGCCAGATAGTATTGAACTGGACCTTTGAAG 838
Db |
212 AATCATCAATGAGAGGCTTTTTCAGCCAGATAGTATTGAACTGGACCTTTGAAG 153
Qy |
839 GAGCAGAGTTTGAAGGTATTCAATATGAAGACATATCTAACTTGTAGTTTGAAGAT 898
Db |
152 GAGCAGAGTTTGAAGGTATTCAATATGAAGACATATCTAACTTGTAGTTTGAAGAT 93
Qy |
899 TTGATAACAACACATTTGAACTGTGAAGCATCAAAATTTGGTGTAGCCAGGCACTGTGTA 958
Db |
92 TTGATAACAACACATTTGAACTGTGAAGCATCAAAATTTGGTGTAGCCAGGCACTGTGTA 33
Qy |
959 ACTTACTGTGTAGGGGATTTGTTTGTATT 990
Db |
32 ACTTACTGTGTAGGGGATTTGTTTGTATT 1

AY412520 1154 bp DNA linear GSS 16-DEC-2003
Mus musculus KIN gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
AY412520 AY412520.1 GI:39768485
GSS.
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 1154)
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejarimal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Perriera, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,
Adams, M.D. and Cargill, M.
TITLE
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
1..1154
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>1154
/genes="KIN"
/locus_tag="HCM4584"

Query Match 45.0%; Score 496.2; DB 10; Length 1154;
Best Local Similarity 99.4%; Pred. No. 2.7e-96;
Matches 498; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

388 GGTACCAAGTGCACCTGAAGCTGCTGGGAGCGGACATCCGGGAAACGGAAGAGTCTTCA 447
Db |
654 GGACCAAGTGCACCTGAAGCTGCTGGGAGCGGACATCCGGGAAACGGAAGAGTCTTCA 713
Qy |
448 CAGAGCTCCGCCAGCTCGAAGAGAGAGAGTCCGCCCTGGATGAGATCATGGAGCTC 507
Db |
714 CAGAGCTCCGCCAGCTCGAAGAGAGAGAGTCCGCCCTGGATGAGATCATGGAGCTC 773
Qy |
508 GAAG 567
Db |
774 GAAG 833
Qy |
568 AAAATTATACGAAG 627
Db |
834 AAAATTATACGAAG 893
Qy |
628 GTGATTGACAGAGTACAGAGCTGTGTTAAAGATGACTCTCGAGACAGCGCTGAAACTG 687
Db |
894 GTGATTGACAGAGTACAGAGCTGTGTTAAAGATGACTCTCGAGACAGCGCTGAAACTG 953
Qy |
688 GACCAGACTCATTTAGAGACAGTCAATTCGCCGCCCGGGGAAAGGGTCTTAGTTTAAAT 747
Db |
954 GACCAGACTCATTTAGAGACAGTCAATTCGCCGCCCGGGGAAAGGGTCTTAGTTTAAAT 1013
Qy |
748 GGAGGCTTACAG 807
Db |
1014 GGAGGCTTACAG 1073


```

RESULT 11
BE225867/c
LOCUS
DEFINITION
    BE225867 513 bp mRNA linear EST 13-MAR-2002
    la20f09.x1 Mouse E10 5 12 5 Pancreas cDNA Library Mus musculus cDNA
    clone IMAGE:5638889 3' similar to TR:060870 O60870 KIN17 PROTEIN.
    ;, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
    Mus musculus (house mouse)
    Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
    Sciurognathi; Muroidea; Muridae; Murinae; Mus.
    1 (bases 1 to 513)
    Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B.,
    Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M.,
    McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y. and
    Bowers, Y.
    WashU-Harvard Pancreas EST Project
    Unpublished (2000)
    Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
    Endocrine Pancreas Consortium
    Harvard University, Howard Hughes Medical Institute
    Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
    MA 02138
    Tel: 617-495-1812
    Fax: 617-495-8557
    Email: dmelton@biohp.harvard.edu
    Library was constructed by Dr. Douglas Melton DNA sequencing by:
    Washington University Genome Sequencing Center For information on
    obtaining a clone please contact: Adimika Meadows
    (meadows@fas.harvard.edu)
    MGI:1847913 This sequence now available from the IMAGE consortium,
    for clone orders contact: info@image.llnl.gov
    Seq primer: -40UP from Gibco
    High quality sequence stop: 384.
FEATURES
    source
    1..513
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="ICR"
    /db_xref="taxon:10090"
    /clone="IMAGE:5638889"
    /sex="Both"
    /tissue_type="Pancreatic Bud"
    /dev_stage="Embryonic day 10.5 and 12.5, mixed"
    /lab_host="DH10B"
    /clone_lib="Mouse E10 5 12 5 Pancreas cDNA Library"
    /notes="Organ: Pancreas; Vector: pSPORT1; Site 1: Not 1;
    Site 2: Sal I; Library constructed using SuperScript
    Plasmid Library kit (Life Technologies). cDNA made by
    oligo-dT priming. Size-selected by column fractionation;
    average insert size 1.47 kb. Primary library, unamplified.
    cDNA Library Preparation: Guolin Chen."
ORIGIN
    Query Match 41.9%; Score 462; DB 2; Length 513;
    Best Local Similarity 97.5%; Pred. No. 5.9e-89;
    Matches 503; Conservative 0; Mismatches 5; Indels 8; Gaps 3;
    QY 550 CAGCCGGGCGTGTGAAATATATACGAAGAGCTTGGGGAATAATCAAGAAG 609
    DB 513 CAGCCGGGCGTGTGTGACATCATACCGAGAGGTT-GGGAGAATATC-CAAGAAG 456
    QY 610 AAAGGGGTGTTAAGGAAGTGATTGACAGGTACACAGCTGTGGTAAAGATGACTCT 669
    DB 455 AAAGGGGTGTTAAGGAAGTGATTGACAGGTACACAGCTGTGGTAAAGATGACTCT 396
    QY 670 GGAGACAGGCTGAACCTGGACCAGACTCATTTAGACAGCTCATTCGGCCCCGGGAAA 729

```

```

DB 395 GGAGACAGGCTGAAACTGGACCAGACTCATTTAGACAGACTCATTCGGCCCCGGGAAA 336
QY 730 AGGGTCTAGTTTTAAATCGAGGCTACAGAGAAATGAAGGCACCTCTCGAATCCATCAAT 789
DB 335 AGGGTCTAGTTTTAAATCGAGGCTACAGAGAAATGAAGGCACCTCTCGAATCCATCAAT 276
QY 790 GAGAAGGCTTTTTCAGCCACAGTAGTCATTGAAACTGGACCTTTGAAAGGAGCGCAGAGTT 849
DB 275 GAGAAGGCTTTTTCAGCCACAGTAGTCATTGAAACTGGACCTTTGAAAGGAGCGCAGAGTT 216
QY 850 GAAGTATTCAATGAAGACATATCTAACTTGTCTTGAATTTGAAAAATTTGATAACAAC 909
DB 215 GAAGTATTCAATGAAGACATATCTAACTTGTCTTGAATTTGAAAAATTTGATAACAAC 156
QY 910 ACATTGAACCTCTGAAGCATCAAAATGGTGTAGCCAGGCACCTGTGTAACCTCTACTGTG 969
DB 155 ACATTGAACCTCTGAAGCATCAAAATGGTGTAGCCAGGCACCTGTGTAACCTCTACTGTG 96
QY 970 TTAGGGGATTTGTTTGTATTAAAAAATAAATCATCTATTAAATCTAGTGAATAG 1029
DB 95 TTAGGGGATTTGTTTGTAT-----AAAAAATCATCTATTAAATCTAGTGAATAG 42
QY 1030 TTGGGTAATTTATATAAATAAATCTATGTTTTTTTAA 1065
DB 41 TTGGGTAATTTATATAAATAAATCTATGTTTTTTTAA 6

```

```

RESULT 12
BE448684
LOCUS
DEFINITION
    BE448684 466 bp mRNA linear EST 25-JUL-2000
    5' similar to TR:060870 O60870 KIN17 PROTEIN. ;, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
    Mus musculus (house mouse)
    Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
    Sciurognathi; Muroidea; Muridae; Murinae; Mus.
    1 (bases 1 to 466)
    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
    National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
    Unpublished (1997)
    Other ESTs: ut48b03.x1
    Contact: Robert Strausberg, Ph.D.
    Email: cgapsb-r@mail.nih.gov
    This clone is available royalty-free through LLNL ; contact the
    IMAGE Consortium (info@image.llnl.gov) for further information.
    MGI:1075249
    Seq primer: -40RP from Gibco
    High quality sequence stop: 465.
FEATURES
    source
    1..466
    /organism="Mus musculus"
    /mol_type="mRNA"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="IMAGE:3331085"
    /sex="male"
    /tissue_type="Spleen"
    /dev_stage="4 weeks"
    /lab_host="DH10B"
    /clone_lib="Soares mouse 3nbms"
    /notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
    polylinker; Site 1: Not 1; Site 2: Eco RI; 1st strand cDNA
    was primed with a Not I - oligo(dT) primer [5',
    TGTTACCAATCTGAAGTGGAGGCGCGCTGTTTTTTTTTTTTTTTTTTT
    3']; double-stranded cDNA was ligated to Eco RI adaptors
    (Pharmacia), digested with Not I and cloned into the Not I
    and Eco RI sites of the modified pT7T3 vector. RNA

```

provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN	Query Match Best Local Similarity Matches 466; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
QY	529 GCACGGACAGACGCTGTGTACAGCCGGGGATCGTTGTGAAATATATACGAAGAAGCTT 588
Db	1 GCACGGACAGACGCTGTGTACAGCCGGGGATCGTTGTGAAATATATACGAAGAAGCTT 60
QY	589 GGGGAGAAATATACAAGAAAGAAAGGGTCGTTAAGGAAGTATTCACAGGTACACAGCT 648
Db	61 GGGGAGAAATATACAAGA--AAAGGGTCGTTAAGGAAGTATTCACAGGTACACAGCT 118
QY	649 GTGGTAAAGATCACTCTCGAGACAGGCTGAAACTGGACACAGCTCATTTAGAGACA 708
Db	119 GTGGTAAAGATCACTCTCGAGACAGGCTGAAACTGGACACAGCTCATTTAGAGACA 178
QY	709 GTCAATTCGGCCCCCGGGAAGGGTCTAGTATTTAAATGGAGGCTACAGAGGAATGAA 768
Db	179 GTCAATTCGGCCCCCGGGAAGGGTCTAGTATTTAAATGGAGGCTACAGAGGAATGNA 238
QY	769 GGCATCTCGAATCAATCAGAAAGGCTTTTCAGCCACGATAGTCAATGAAACTGGA 828
Db	239 GGCATCTCGAATCAATCAGAAAGGCTTTTCAGCCACGATAGTCAATGAAACTGGA 298
QY	829 CTTTGAAGGACGAGAGTCAAGGTATTCATATGAGACATATCTAAACTCTCTCA 888
Db	299 CTTTGAAGGACGAGAGTCAAGGTATTCATATGAGACATATCTAAACTCTCTCA 358
QY	889 GTTGAATAATTTGATAACACACATTTGAACTGTGAAAGCATCAAAATTTGGTTAGCCAAG 948
Db	359 GTTGAATAATTTGATAACACACATTTGAACTGTGAAAGCATCAAAATTTGGTTAGCCAAG 418
QY	949 GCATGTGTAACTCTACTGTGTAGGGGATTTGTTTGTATTAATAA 996
Db	419 GCATGTGTAACTCTACTGTGTAGGGGATTTGTTTGTATTAATAA 466

RESULT 13	
AY609991	1543 bp mRNA linear HTC 31-JAN-2005
LOCUS	
DEFINITION	Sus scrofa clone Clu_5302.scr.msk.pl.Contig4, mRNA sequence.
ACCESSION	AY609991
VERSION	AY609991.1 GI:52351561
KEYWORDS	HTC.
SOURCE	Sus scrofa (pig)
ORGANISM	Sus scrofa
REFERENCE	Jorgensen,F.G., Hobolth,A., Hornshoj,H., Bendixen,C., Fredholm,M. and Schierup,M.H.
AUTHORS	Comparative analysis of protein coding sequences from human, mouse and the domesticated pig
TITLE	(er) BMC Biol. 3 (1), 2 (2005)
JOURNAL	15679890
PUBMED	2 (bases 1 to 1543)
REFERENCE	Hornshoj,H., Bendixen,C. and Panitz,F.
AUTHORS	Direct Submission
TITLE	Submitted (28-APR-2004) Animal Breeding and Genetics, Danish
JOURNAL	Institute of Agricultural Sciences, Research Centre Foulum, Postbox
	50, Tjele DK-8830, Denmark
FEATURES	Location/Qualifiers
source	1..1543
	/organism="Sus scrofa"
	/mol_type="mRNA"
	/db_xref="taxon:9823"
	/clone="Clu_5302.scr.msk.pl.Contig4"

misc_feature 1..1543
/note="similar to NM_012311.2 Homo sapiens KIN, antigenic determinant of recA protein homolog (mouse) (KIN)"

ORIGIN	Query Match Best Local Similarity Matches 871; Conservative 0; Mismatches 160; Indels 296; Gaps 3;
QY	1 ATGGSCAAGTCGGATTTTCTCAGGCCCAAGGCCATCGCCAATAGAAATTAAGTCCAAAGGG 60
Db	36 ATGGSCAAGTCGGATTTTCTACCCGAAGGCCATCGCCAACAGGATCAAAATCCAAGGG 95
QY	61 CTCAGAAAGCTTCGCTGTGATCTGCAGATGTGCCAAAGCAATGCCGCGACAGAAATGCG 120
Db	96 CTTCAGAAAGCTCGCGTGGTATTGCCAGATGTGCCAAGAGCAGTGCCTGGACAGAAATGCG 155
QY	121 TTTAAGTGTCACTGTATGCTGAATCTCATCAAGACAACTGTTGCTGCTTCAGAAAAC 180
Db	156 TTTAAGTGTCACTGTATGCTCGAATCGCATCAGACAACTGTTGCTGCTTCTGAAAT 215
QY	181 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTTCCGAAATGACTTTCTGGAATCTCTG 240
Db	216 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTTCCGAAATGACTTTCTAGAACTCTC 275
QY	241 AGGCGACGCTTTGGCACTAAAGGGTCCAAACAACATTTGCTACAATGAATACATCAGC 300
Db	276 AGGAGACGCTTTGGCACCAAGAGGGTTCAACAACAACATCGTCTACAACGAGTACATCAGC 335
QY	301 CACCGAGACACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAG 360
Db	336 CACCGAGACACATCCACATGAACGCTACCCAGTGGGAGACGCTGACCGAATTTCCACAG 395
QY	361 TGGCTGGCGAGAGGGCTTGTGTA 387
Db	396 TGGCTGGCGAGAGGGCTTGTGC 455
QY	388 387
Db	456 TACATAGATAGGGACCCCGAACAACATCCCGCAGCACTGGAACTAGAGAAAAGAGAG 515
QY	388 387
Db	516 CAGGACTTTGGACGATGAAGAAAAAACTCCCAAAATTTATTCAGAAACAAGTGAGACGAGT 575
QY	388 387
Db	576 CTGGAAGGNAAGGAGCAGGAGGCCCTTATTTTACGAGTTAAGCAGAGAAAAATGAAGAA 635
QY	388 387
Db	636 GAAAAAGTGACATTTAAATTTGAAACAAAGAGCAGTGTAGTTCAGCAGCAGCATCTTTCCAAA 695
QY	388 435
Db	636 TCAAGTTCTTTGGGACCAAGTGCCCTGAAGACAGATGGGTAGCAGACGCTCGGTGAAACA 755
QY	436 AAAGAGCTTTCACAGAGCTCCGCCACGAGCTGC-----GAAGAAGAAGAAGTCGGCCCTG 489
Db	756 AAGGAATCTTCCAGAGCTCAGTCAGTCAAAAGAAAAGNAGNAAAAGAGTCTCGGCTC 815
QY	490 GATGAGATCATGGAGCTCGAAGAGAGAAAAGAAAGAGCCGACGAGACAGCCCTGGTTA 549
Db	816 GACGAAATCATGGAGATTTGAAGAGAGAGAAAGAGAACTGCCCGACGGATCAGCTGCTC 875
QY	550 CAGCCGGGATCGTTGTGAAAATTTATACGAGAGAGCTTTGGGAGAAAATATCACAGAGAG 609
Db	876 CAGCTGAAATTTATAGTGAATAATTTATACCAAAAACCTGGAGAGAAATATCATAGAAA 935
QY	610 AAAGGGTCTGTTAAGGAAAGTGAATTCACAGGTACACAGCTGTGGTAAAGATGACTCTCT 669
Db	936 AAGGCGATTTGTTAAGGAAGTAATTGACAAAATACACAGCTGTTGTTAAGATGATTGACTCT 995

REFERENCE
AUTHORS

1 (bases 1 to 447)
Akinura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiranoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
Matsumura,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
Okazaki,Y., Okido,T., Saito,K., Sakai,C., Sakai,K., Sakazume,N.,
Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yaeunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akinura,T., et al.
2001)

JOURNAL
COMMENT

Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES
source

Location/Qualifiers
1..447
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="G030102N09"
/clone_lib="RIKEN full-length enriched, pooled tissues,
pituitary, etc."
/note="pooled tissues ; (tissue type=pituitary gland,
dev_stage=adult, sex=male), (tissue type=thymus,
dev_stage=adult, sex=male), (tissue type=forelimb,
dev_stage=13 days embryo, sex=mixed), (tissue_type=testis,
dev_stage=13 days embryo, sex=male), (tissue_type=olfactory brain, dev_stage=adult, sex=male)"

ORIGIN

Query Match 39.3%; Score 433.4; DB 2; Length 447;
Best Local Similarity 99.6%; Pred. No. 8.5e-83;
Matches 445; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY	624	GGAAAGTGAATTGACAGTACACAGCTGTGGTAAAGATGACTCTCTGGAGACAGGCTGAA	683
Db	1	GGAAAGTGAATTGAAAGGTACACAGCTGTGGTAAAGATGACTCTCTGGAGACAGGCTGAA	60
QY	684	ACTGGACAGACTCATTAGACAGCTCATTCCGCCCGGGGAAAGGGTTCTAGTTT	743
Db	61	ACTGGACAGACTCATTATAGACAGCTCATTCCGCCCGGGGAAAGGGTTCTAGTTT	120
QY	744	AAATGGAGGCTACAGAGGAAATGAAGGCACTCTCGAATCATCAATGAGAAGGCTTTTC	803
Db	121	AAATGGAGGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCAATGAGAAGGCTTTTC	180

QY	804	AGCCACGATAGTCATTGAAACTGGACCTTTGAAAGGACGACAGGTTGAAGGTATTCAATA	863
Db	181	AGCCACGATAGTCATTGAAACTGGACCTTTGAAAGGACGACAGGTTGAAGGTATTCAATA	240
QY	864	TGAAGACATATCTAAACTTTCCTTGAGTTTGAATAATTTTGATAACACACATTTGAAACTGTG	923
Db	241	TGAAGACATATCTAAACTTTCCTTGAGTTTGAATAATTTTGATAACACACATTTGAAACTGTG	300
QY	924	AAGCATCAAAATTTGGTGTAGCCAAAGGCACTGTGTAACCTCTACTGTGTTAGGGGATTTGTT	983
Db	301	AAGCATCAAAATTTGGTGTAGCCAAAGGCACTGTGTAACCTCTACTGTGTTAGGGGATTTGTT	360
QY	984	TTGTATTAAAAAATAAATAATCATCTATTAAATACTAGTGAATAGTTGGGTAAATTTAT	1043
Db	361	TTGTATT-AAAAAATAAATAATCATCTATTAAATACTAGTGAATAGTTGGGTAAATTTAT	419
QY	1044	AATAAAATCTATGTTTTTTTTTTAAAGTGT	1070
Db	420	AATAAAATCTATGTTTTTTTTTTAAAGTGT	446

Search completed: November 27, 2005, 00:57:45
Job time : 6166.01 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:02 ; Search time 228.214 Seconds
(without alignments)
8583.479 Million cell updates/sec

Title: US-09-555-529-2
Perfect score: 1102
Sequence: 1 atggcgaagtcggatttctt.....aaaaaaaaaaaaaaaaaaaa 1102

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/1 COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5 COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/1/ina/H COMB.seq: *
6: /cgn2_6/ptodata/1/ina/PTUS COMB.seq: *
7: /cgn2_6/ptodata/1/ina/PP COMB.seq: *
8: /cgn2_6/ptodata/1/ina/RE COMB.seq: *
9: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	118.8	10.8	538	3	US-09-270-767-379, App
c 2	118.8	10.8	538	3	US-09-270-767-15661
c 3	90	8.2	23645	3	US-09-949-016-13916
c 4	69.4	6.3	7218	2	US-08-232-463-14
5	61.8	5.6	468	3	US-09-270-767-9580
6	61.8	5.6	468	3	US-09-270-767-24862
7	60	5.4	825	3	US-09-248-796A-5938
8	59	5.4	2229	3	US-08-207-954-6
9	59	5.4	2510	2	US-08-888-982A-42
10	59	5.4	2510	3	US-09-462-261-42
11	59	5.4	2510	3	US-09-506-073-89
12	59	5.4	2510	3	US-10-057-550C-67
13	54.8	5.0	3680	3	US-09-647-390-15
14	54.6	5.0	332	3	US-09-621-976-16031
15	53.2	4.8	1752	3	US-09-292-225-14
16	53.2	4.8	1752	3	US-09-292-225-16
17	53.2	4.8	1753	3	US-09-149-476-56
18	52.8	4.8	1039	3	US-09-902-540-1280
19	52.8	4.8	2797	3	US-09-482-273-74
20	52.4	4.8	247299	3	US-09-949-016-17590
21	52.2	4.7	1553	3	US-09-280-116-10
22	51.6	4.7	1976	3	US-09-920-759-10
23	51.2	4.6	785	3	US-09-205-258-73
24	51.2	4.6	785	3	US-10-004-860-73

25	51.2	4.6	41736	3	US-09-949-016-17091	Sequence 17091, A
26	51	4.6	306	3	US-09-621-976-16035	Sequence 16035, A
27	51	4.6	326	3	US-09-621-976-16024	Sequence 16024, A
28	51	4.6	329	3	US-09-621-976-16012	Sequence 16012, A
29	51	4.6	332	3	US-09-621-976-16053	Sequence 16053, A
30	51	4.6	333	3	US-09-621-976-16032	Sequence 16032, A
31	51	4.6	333	3	US-09-621-976-16045	Sequence 16045, A
32	51	4.6	334	3	US-09-621-976-16044	Sequence 16044, A
33	51	4.6	336	3	US-09-621-976-16013	Sequence 16013, A
34	51	4.6	347	3	US-09-621-976-16026	Sequence 16026, A
35	51	4.6	357	3	US-09-621-976-16058	Sequence 16058, A
36	51	4.6	359	3	US-09-621-976-16008	Sequence 16008, A
37	51	4.6	359	3	US-09-621-976-16019	Sequence 16019, A
38	51	4.6	362	3	US-09-621-976-16010	Sequence 16010, A
39	51	4.6	365	3	US-09-621-976-16042	Sequence 16042, A
40	51	4.6	371	3	US-09-621-976-16048	Sequence 16048, A
41	50.6	4.6	305	3	US-09-621-976-16020	Sequence 16020, A
42	50.6	4.6	306	3	US-09-621-976-16057	Sequence 16057, A
43	50.6	4.6	327	3	US-09-621-976-16018	Sequence 16018, A
44	50.6	4.6	387	3	US-10-125-258-104	Sequence 104, App
45	50.2	4.6	332	3	US-09-621-976-16050	Sequence 16050, A

ALIGNMENTS

RESULT 1

US-09-270-767-379/c
; Sequence 379, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 379
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-379

Query Match 10.8%; Score 118.8; DB 3; Length 538;
Best Local Similarity 56.3%; Pred. No. 1.6e-18;
Matches 222; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

Qy	496	ATCATGGAGCTCGAAGAGGAAAAAGGACCGCACGACGCGCTGGTTACAGCGG	555
Db	522	ATCAAGCAGGAGGAAAGCAAAAGGAGCGTGCCCAACCGCAAGACTACTGGCTGCAAG	463
Qy	556	GGATCGTTGTGAAAAATTATAACGAGAGAGCTTTGGGAGAAATATCACAAGAGAAAGGG	615
Db	462	GGTATCGTGTGCAATTTATTTCCAAATCCATCGGCGGAAAAGTTCTTCAACACAAAAAGCG	403
Qy	616	GTCTTAAAGAAAGTATTGACAGGTACACAGCTGTGGTAAAGATGACTCTGGAGAC	675
Db	402	GTGTCTCGGACGTAAATTCAGATATCAGGGCAAAATCAAGTTCTTTGGAGACTGGGGAA	343
Qy	676	AGCTGAAACTGGACACAGACTCATTTAGACAGACTCATTCGGCGCCCGGGGAAAAGGTTT	735
Db	342	AACTAAAGTGAATCAAGTCAATTTGGAGACGGTAATCCCGCTTTGGACAAAGCCTGTC	283
Qy	736	CTAGTTTTTAATGGAGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCAATGAGAG	795
Db	282	ATGTTGGTTAATGGCGTTATCGGGATCCGAGGCTCTGCTAAGGAAACTGGACGAGCGC	223
Qy	796	CTTTTTCAGCCAGTAGTCAATTTGAACCTGGACCTTTGAAGGACCCAGAGTTGAAGGT	855
Db	222	AGATATTTCAGTCAGCGTGGAAATATTGACAGCGTCTCTCAAGGCGAGATTTGTAGACAC	163
Qy	856	ATTCATATGAGACATATCTAAACTTGTGTTGAG	889

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13916
; LENGTH: 23645
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13916

Query Match      8.2%; Score 90; DB 3; Length 23645;
Best Local Similarity 86.8%; Pred. No. 3.5e-11;
Matches 99; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ATGGCCAAAGTCTGGATTTCTGAGCCCAAGCCCAATGCAATTAAGTCCAAAGG 60
DB 1763 ATGGGAAAGTCGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAAAGGG 1704
QY 61 CTCACAAAGCTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAG 114
DB 1703 CTGCAGAAGCTACGCTGGTATTGCCAGATGTGCCAGAAGCACTGCCGCGACGAG 1650

RESULT 4
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match      6.3%; Score 69.4; DB 2; Length 7218;
Best Local Similarity 2.6%; Pred. No. 1.7e-06;
Matches 10; Conservative 236; Mismatches 137; Indels 0; Gaps 0;
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13916
; LENGTH: 23645
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13916

Query Match      8.2%; Score 90; DB 3; Length 23645;
Best Local Similarity 86.8%; Pred. No. 3.5e-11;
Matches 99; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ATGGCCAAAGTCTGGATTTCTGAGCCCAAGCCCAATGCAATTAAGTCCAAAGG 60
DB 1763 ATGGGAAAGTCGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAAAGGG 1704
QY 61 CTCACAAAGCTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGCGACGAG 114
DB 1703 CTGCAGAAGCTACGCTGGTATTGCCAGATGTGCCAGAAGCACTGCCGCGACGAG 1650

RESULT 4
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match      6.3%; Score 69.4; DB 2; Length 7218;
Best Local Similarity 2.6%; Pred. No. 1.7e-06;
Matches 10; Conservative 236; Mismatches 137; Indels 0; Gaps 0;
```

```
QY 395 GTCCACTGAAGCTGCTGGGAGCGCAGCATCCGGAAACGGAAGAGTCTTCACAGAGCT 454
Db 1438 GTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1379
QY 455 CCGCCAGCTCGAAGAAAGAACTCGGCCCTCGATGAGATCATGGAGCTCGAAGAG 514
Db 1378 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1319
QY 515 AAAAGAAAAGGACCGCAGCAGACGCTGTTTACAGCGGGGATCGTTTGAAATTA 574
Db 1318 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1259
QY 575 TAACGAAGAAGCTGGGGAGAAATATACAGAAGAAAGGGTCTTAAGGAAGTCAATG 634
Db 1258 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1199
QY 635 ACAGGTACACAGCTGTGTAAGATGACTCTGGAGACAGCTGAACTGACACGAGA 694
Db 1198 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1139
QY 695 CTCATTTAGAGACAGTCATTCGGGCCCGGGGAAAGGGTCTAGTTTAAATGGAGCT 754
Db 1138 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1079
QY 755 ACAGGAAGATGAAGCACTCTC 777
Db 1078 RRRRRRRRRRRATCGCAAGCTC 1056
```

```
RESULT 5
US-09-270-767-9580
; Sequence 9580, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9580
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-9580
```

```
Query Match 5.6%; Score 61.8; DB 3; Length 468;
Best Local Similarity 64.1%; Pred. No. 4.4e-05;
Matches 93; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 958 AACTCTACTGTGTAGGGGATTTGTTTGTATTAAAAAATAAATCATCTATTTAAAT 1017
Db 219 AGCTTTAGCTTTTATCAGGATATCATATATTAAGAAGAAATTTATGCTCTTTAAT 278
QY 1018 ACTAGTGAATGTGGGTAAATTTATATAAATCTATGTTTAAAGTGTAAAAA 1077
Db 279 TTATATAAAATTTTAAATAATTCATAATCAATATATATGTTGTTAGTTTCATGAACA 338
QY 1078 AAAAAAATAAATAAATAAATAA 1102
Db 339 CAATGAATAAATAAATAAATAA 363
```

```
RESULT 6
US-09-270-767-24862
; Sequence 24862, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
```

```
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24862
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-24862

Query Match 5.6%; Score 61.8; DB 3; Length 468;
Best Local Similarity 64.1%; Pred. No. 4.4e-05;
Matches 93; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 958 AACTCTACTGTGTAGGGGATTTGTTTGTATTAAAAAATAAATCATCTATTTAAAT 1017
Db 219 AGCTTTAGCTTTTATCAGGATATCATATTAAGAAGAAATTTATGCTCTTTAAT 278
QY 1018 ACTAGTGAATGTGGGTAAATTTATATAAATCTATGTTTAAAGTGTAAAAA 1077
Db 279 TTATATAAAATTTTAAATAATTCATAATCAATATATATGTTGTTAGTTTCATGAACA 338
QY 1078 AAAAAAATAAATAAATAAATAA 1102
Db 339 CAATGAATAAATAAATAAATAA 363
```

```
RESULT 7
US-09-248-796A-5938
; Sequence 5938, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5938
; LENGTH: 825
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5938
```

```
Query Match 5.4%; Score 60; DB 3; Length 825;
Best Local Similarity 49.0%; Pred. No. 0.00014;
Matches 193; Conservative 0; Mismatches 195; Indels 6; Gaps 1;

QY 1 ATGGGCAAGTCGAGTTTCTGAGCCCAAGGCCATCGCCCAATAGAATTAAGTCCAAAGG 60
Db 106 ATGGCAAGGCGAGAGTTTGGAAACAGCAAAATATCAATCTAAAAAGCTACGAGCGGCGA 165
QY 61 CTCAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCCCGCAGAGATGGC 120
Db 166 TTACAAAAATTTGAAGTTCTATTCGCAATTTGTGTTCTTAAACAGTGTGAGATTCATGGG 225
QY 121 TTTAAGTGTCTGTATGTCTGAATCTCATCAAGAACAACTGT-----TGCTGGGTTCA 174
Db 226 TTTAAAAACCATCTATCTTCACCGCTGCACATAAAGAGGTTCCGAAATACATGAATCT 285
QY 175 GAAAAACCTCAGCAGTTTATGGAATTTTTCAGAGGAATTCGAAATGATCTTCTGGAA 234
Db 286 GGTGATAGCTCCAAAGCTAATAGAAACCTACTCTACAAAATTTCCAAGATAAGTTTATCAAG 345
QY 235 CTTCTGAGGCGAGCTTTGGCACTAAAGGGTCCACAAACAACTTGTCTCAATGAATAC 294
Db 346 TTGCTCCGAATCAACCATGGAAGTAAATTTTATCAATGCCCAATAGTTTCTACCAAGATAT 405
QY 295 ATCAGCCACCGAGAGCAGACATCCACATGAACGCTACCCAGTGGGAGAGACTGACCGACTTT 354
```



```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: Pentium
; OPERATING SYSTEM: Windows 95
; SOFTWARE: WORDPERFECT 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/462,261
; FILING DATE: 01-Mar-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/756,806
; FILING DATE: No. 6391636ember 26, 1996
; APPLICATION NUMBER: PCT/US95/07111
; FILING DATE: May 31, 1995
; APPLICATION NUMBER: 08/250,856
; FILING DATE: May 31, 1994
; APPLICATION NUMBER: 08/888,982
; FILING DATE: July 7, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane Massey Licata
; REGISTRATION NUMBER: 32,257
; REFERENCE/DOCKET NUMBER: ISPH-0312
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 779-2400
; TELEFAX: (609) 810-1454
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2510
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-462-261-42
```

```
Query Match 5.4%; Score 59; DB 3; Length 2510;
Best Local Similarity 69.6%; Pred. No. 0.00035;
Matches 80; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 988 ATTAAAAAATAAATCATCTATTAAATAGTGAATAGTTGGTAAATTTATAATA 1047
DB 2386 AGTAGCAACAAAGGAAATAAATGAACATATGTTGCTTATATGTTAAATGAATAAAA 2445
QY 1048 AAATCTATGTTTTTTTAAAGTGTAATAAATAAATAAATAAATAAATAAATAA 1102
DB 2446 TACTCTCTTTTTTTTAAAGTGGAATAAATAAATAAATAAATAAATAAATAA 2500
```

```
RESULT 11
US-09-506-073-89
; Sequence 89, Application US/09506073
; Patent No. 6410518
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; TITLE OF INVENTION: Antisense Oligonucleotide Modulation of raf Gene Expression
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/506,073
; CURRENT FILING DATE: 2000-02-18
; EARLIER APPLICATION NUMBER: US 09/143,214
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: PCT/US98/13961
; EARLIER FILING DATE: 1998-07-06
; EARLIER APPLICATION NUMBER: US 08/888,982
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US 08/756,806
; EARLIER FILING DATE: 1996-11-26
; EARLIER APPLICATION NUMBER: PCT/US95/07111
; EARLIER FILING DATE: 1995-05-31
; EARLIER APPLICATION NUMBER: US 08/250,856
; EARLIER FILING DATE: 1994-05-31
; NUMBER OF SEQ ID NOS: 130
; SEQ ID NO 89
; LENGTH: 2510
```

```
;
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; US-09-506-073-89
Query Match 5.4%; Score 59; DB 3; Length 2510;
Best Local Similarity 69.6%; Pred. No. 0.00035;
Matches 80; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 988 ATTAAAAAATAAATCATCTATTAAATAGTGAATAGTTGGTAAATTTATAATA 1047
DB 2386 AGTAGCAACAAAGGAAATAAATGAACATATGTTGCTTATATGTTAAATGAATAAAA 2445
QY 1048 AAATCTATGTTTTTTTAAAGTGTAATAAATAAATAAATAAATAAATAAATAA 1102
DB 2446 TACTCTCTTTTTTTTAAAGTGGAATAAATAAATAAATAAATAAATAAATAA 2500

RESULT 12
US-10-057-550C-67
; Sequence 67, Application US/10057550C
; Patent No. 6806258
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P
; TITLE OF INVENTION: Antisense Oligonucleotide Modulation of raf Gene Expression
; FILE REFERENCE: ISPH-0625
; CURRENT APPLICATION NUMBER: US/10/057,550C
; CURRENT FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: US 09/506,073
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: US 09/143,214
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 08/756,806
; PRIOR FILING DATE: 1996-11-26
; PRIOR APPLICATION NUMBER: PCT/US95/07111
; PRIOR FILING DATE: 1995-05-31
; PRIOR APPLICATION NUMBER: US 08/250,856
; PRIOR FILING DATE: 1994-05-31
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 2510
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-057-550C-67
```

```
Query Match 5.4%; Score 59; DB 3; Length 2510;
Best Local Similarity 69.6%; Pred. No. 0.00035;
Matches 80; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 988 ATTAAAAAATAAATCATCTATTAAATAGTGAATAGTTGGTAAATTTATAATA 1047
DB 2386 AGTAGCAACAAAGGAAATAAATGAACATATGTTGCTTATATGTTAAATGAATAAAA 2445
QY 1048 AAATCTATGTTTTTTTAAAGTGTAATAAATAAATAAATAAATAAATAAATAA 1102
DB 2446 TACTCTCTTTTTTTTAAAGTGGAATAAATAAATAAATAAATAAATAAATAA 2500

RESULT 13
US-09-647-390-15
; Sequence 15, Application US/09647390
; Patent No. 6465636
; GENERAL INFORMATION:
; APPLICANT: Stuiver, Maarten
; APPLICANT: Cuisters, Jerome
; APPLICANT: Simons, Lambertus
; TITLE OF INVENTION: Pathogen-Inducible Promoter
; FILE REFERENCE: MOG 57707/UST
; CURRENT APPLICATION NUMBER: US/09/647,390
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: EP 98201024.1
; PRIOR FILING DATE: 1998-04-01
```

```
; PRIOR APPLICATION NUMBER: PCT/EP99/02178
; PRIOR FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 15
; LENGTH: 3680
; TYPE: DNA
; ORGANISM: Helianthus annuus
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1889)
; NAME/KEY: CDS
; LOCATION: (1890)..(3503)
US-09-647-390-15

Query Match          5.0%; Score 54.8; DB 3; Length 3680;
Best Local Similarity 54.5%; Pred. No. 0.0039;
Matches 110; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 895 AAATTGTGATAACACACATGAACTGTGAAGCATCAAAATGGTGTAGCCAAAGGCACGTG 954
Db ||||| || || || || || || || || || || || || || || || || || || ||
697 AATTTTTTTTAAATAAATTAATGAAATTTTAAAGAAAGATCATTTTCTAAAGATCCGTAGCG 756
QY 955 TGTAACTCTACTGCTGTGTAGGGGATTTGTTTGTATTAAAAAATAATCATCTATTTA 1014
Db ||||| || || || || || || || || || || || || || || || || || || ||
757 AGTAAAGTTATGATGTTGTCTAACTTTTATGTTCTTATTCATCTGTTTAAATATA 816
QY 1015 AATACTAGTGAATAGTGGGTAAATTTATAATAAAATCTATGTTTTTTTAAAGTGTAAAA 1074
Db || || || || || || || || || || || || || || || || || || || || ||
817 TAAAAAGATAAGGAGTGTGTAAACAAATATAAAGAGTGTGTTAAAGGTAAAAAATAA 876
QY 1075 AAAAAAAAAAAAAAAAAAAAAA 1096
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
877 AAAAAAAAAAAAAAAAAAAAAA 898

RESULT 14
US-09-621-976-16031
; Sequence 16031, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S. J.Y.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 16031
; LENGTH: 332
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-16031

Query Match          5.0%; Score 54.6; DB 3; Length 332;
Best Local Similarity 65.2%; Pred. No. 0.002;
Matches 75; Conservative 3; Mismatches 37; Indels 0; Gaps 0;

QY 988 ATTAAAAAATAAATAATCATCTATTAAATACACTAGTGAATAGTTGGTAAATTTATAATA 1047
Db ||: ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
192 ATGAAATAGAAAAAATAATTTTCGATAGAAAAATAAATAGAAAAATAGAAAAAATTTAAAAA 251
QY 1048 AAATCTATGTTTTTTTTTAAAGTGTAAAAAATAAATAAATAAATAAATAAATAAATAA 1102
Db ||||| : || || || || || || || || || || || || || || || || || || || || ||
252 CAACCMAGCCTCATCTATGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 306

RESULT 15
US-09-292-225-14
; Sequence 14, Application US/09292225
; Patent No. 6455686
```

```
; GENERAL INFORMATION:
; APPLICANT: McCall, Catherine A.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Weber, Eric R.
; TITLE OF INVENTION: NOVEL DERMATOPHAGOIDES NUCLEIC ACID MOLECULES, PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: AL-2-C3
; CURRENT APPLICATION NUMBER: US/09/292,225
; CURRENT FILING DATE: 1999-04-15
; EARLIER APPLICATION NUMBER: 60/098,909
; EARLIER FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: 60/085,295
; EARLIER FILING DATE: 1998-05-13
; EARLIER APPLICATION NUMBER: 60/098,565
; EARLIER FILING DATE: 1998-04-17
; EARLIER APPLICATION NUMBER: 09/062,013
; EARLIER FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 14
; LENGTH: 1752
; TYPE: DNA
; ORGANISM: Dermatophagoides farinae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1665)
US-09-292-225-14

Query Match          4.8%; Score 53.2; DB 3; Length 1752;
Best Local Similarity 61.6%; Pred. No. 0.0072;
Matches 85; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 965 CTGTGTTAGGGGATTTGTTTGTATTAAAAAATAATCATCTATTTAATAATACTAGTG 1024
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1614 CTGTCCACCGGCGCACTATTGCGTCAAGAAAAATTGACTTGTATAGCGAATAATTCTG 1673
QY 1025 AATAGTTGGGTAAATTTTATAATAAAATCTATGTTTTTTTTTAAGTGTAAAAAATAA 1084
Db || || || || || || || || || || || || || || || || || || || || ||
1674 AAAAAAATAATCAATTAATAAATTTAAAAATTCATTTTTTAATATGAAAAAATTCAAAAA 1733
QY 1085 AAAAAAAAAAAAAAAAAAAAAA 1102
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1734 AAAAAAAAAAAAAAAAAAAAAA 1751

Search completed: November 27, 2005, 01:10:38
Job time : 230.214 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 ; Search time 1270.42 Seconds
(without alignments)
7173.088 Million cell updates/sec

Title: US-09-555-529-2
Perfect score: 1102
Sequence: 1 atggcgcaagtcggattttct.....aaaaaaaaaaaaaaaaaaaa 1102

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.Main:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	309	28.0	461	7	US-10-242-535A-29177 Sequence 29177, A
2	309	28.0	461	7	US-10-085-783A-29177 Sequence 29177, A
3	269.2	24.4	591	5	US-10-106-698-1187 Sequence 1187, Ap
4	172.4	15.6	538	7	US-10-437-963-89433 Sequence 89433, A
5	169.2	15.4	2046	8	US-10-425-115-157125 Sequence 157125, Ap
6	168.8	15.3	1614	7	US-10-425-114-22975 Sequence 22975, Ap
7	168.8	15.3	1713	8	US-10-425-114-22614 Sequence 22614, A
8	167.6	15.2	1713	8	US-10-739-930-2955 Sequence 2955, Ap
9	165	15.0	777	7	US-10-424-599-25484 Sequence 25484, A
10	163	14.8	1241	10	US-11-097-143-41039 Sequence 41039, A
11	163	14.8	3313	10	US-11-097-143-41038 Sequence 41038, A
12	163	14.8	4394	10	US-11-097-143-41017 Sequence 41017, A
13	163	14.8	4582	10	US-11-097-143-24532 Sequence 24532, A
14	150.4	13.6	584	7	US-10-767-701-25587 Sequence 25587, A
15	130.8	11.9	549	3	US-09-991-936-1424 Sequence 1424, Ap
16	130.8	11.9	549	3	US-10-978-245-1424 Sequence 1424, Ap
17	95.4	8.7	300	9	US-10-779-543-7646 Sequence 7646, Ap
18	93.6	8.5	270	3	US-09-294-093B-3089 Sequence 3089, A
19	91.2	8.3	739	7	US-10-424-599-53324 Sequence 53324, A
20	81.2	7.4	516	4	US-09-925-065A-549989 Sequence 549989, Ap
21	67.2	6.1	521	8	US-10-425-115-116286 Sequence 116286, A
22	66.2	6.0	375	8	US-10-357-930-44930 Sequence 44930, A
23	65	5.9	65	3	US-09-908-975-24642 Sequence 24642, A

24	63.6	5.8	616	7	US-10-424-599-53326	Sequence 53326, A
c 25	63	5.7	396	7	US-10-021-323-17801	Sequence 17801, A
26	62.6	5.7	552	7	US-10-021-323-93	Sequence 93, Appl
27	61.8	5.6	320	8	US-10-425-115-169025	Sequence 169025, Ap
c 28	61.2	5.6	723	3	US-09-814-353-4726	Sequence 4726, Ap
c 29	61.2	5.6	723	3	US-09-814-353-11024	Sequence 11024, A
c 30	59.6	5.4	465	8	US-10-425-115-170017	Sequence 170017, A
c 31	59.4	5.4	390	3	US-09-814-353-17808	Sequence 17808, A
c 32	59.4	5.4	801	7	US-10-437-963-49284	Sequence 49284, A
c 33	59.2	5.4	327	8	US-10-357-930-8158	Sequence 8158, Ap
34	59	5.4	2229	6	US-10-440-341-6	Sequence 6, Appli
35	59	5.4	2510	3	US-09-997-722-203	Sequence 203, App
36	59	5.4	2510	5	US-10-057-550-89	Sequence 89, Appl
37	59	5.4	2510	6	US-10-173-225B-67	Sequence 67, Appl
38	59	5.4	2510	7	US-10-444-190-1	Sequence 1, Appli
39	59	5.4	2510	8	US-10-834-665-1	Sequence 1, Appli
c 40	58.6	5.3	284	3	US-09-814-353-17553	Sequence 17553, A
c 41	58.4	5.3	510	3	US-09-814-353-18805	Sequence 18805, A
c 42	58.2	5.3	453	3	US-09-814-353-5797	Sequence 5797, Ap
c 43	58.2	5.3	453	3	US-09-814-353-12080	Sequence 12080, A
c 44	58.2	5.3	1184	8	US-10-425-115-20065	Sequence 20065, A
45	58	5.3	277	3	US-09-960-352-12673	Sequence 12673, A

ALIGNMENTS

RESULT 1
US-10-242-535A-29177
; Sequence 29177, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29177
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-29177
Query Match 28.0%; Score 309; DB 7; Length 461;
Best Local Similarity 90.5%; Pred. No. 1.6e-56;
Matches 352; Conservative 0; Mismatches 35; Indels 2; Gaps 2;
Qy 1 ATGGCGAAGTCGGAATTTCTGAGCCGCCCAAGCCCATGCCCAATAGATTAAGTCCAAAGGG 60
Db 8 ATGGCGAAGTCGGAATTTCTTACTCCCAA-GCTATGCCCAACAGGATCAAGTCCAAAGGG 66
Qy 61 CTCAGAGGTTGCTGGTACTGCCAGATGTCGCAAGCAATGCGCCGACGAGAAATGGC 120
Db 67 CTCAGAGGTTGCTGGTACTGCCAGATGTCGCAAGCAATGCGCCGACGAGAAATGGC 126
Qy 121 TTTAAGTGTCACGTGTATGCTGAATCTCATCAAGACAACCTGTTGCTGGCTTCAGAAAAC 180
Db 127 TTTAAGTGTCATGTATGTCGGAATCTCACCAGAGACAATATTGCTGGCTTCAGAAAAT 186
Qy 181 CCTCAGAGTTATGGAATATTTTTCAGAGAAATTCGAGAAATGACTTTCTGGAACCTTCG 240
Db 187 CCTCAGAGTTATGGAATATTTTTCAGAGAAATTCGAGAAATGACTTTCTGGAACCTTC 246

;; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

;; FILE REFERENCE: 38-21(53221)B

;; CURRENT APPLICATION NUMBER: US/10/437,963

;; CURRENT FILING DATE: 2003-05-14

;; NUMBER OF SEQ ID NOS: 204966

;; SEQ ID NO 89433

;; LENGTH: 538

;; TYPE: DNA

;; ORGANISM: Oryza sativa

;; FEATURE:

;; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1

US-10-437-963-89433

Query Match 15.6%; Score 172.4; DB 7; Length 538;
Best Local Similarity 65.1%; Pred. No. 4.8e-27;
Matches 254; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 1 ATGGGCAAGTCGGATTTCTGAGCCCAAGGCCATCGCCAATAGAATTAAAGTCCAAAGGG 60
DB |||||
QY 124 ATGGGGAAGCACGAGTTCCTGAGCCGGAAGGCCATCGGNAAGGATCAAGGCGAAGGGG 183
DB |||||
QY 61 CTCGAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCCGCGACGAGATGGC 120
DB |||||
QY 184 CTGCAAGCTGCGGTGGTACTGCCAGATGTGCCAGAGCAGTGGCCGACGAGATGGC 243
DB |||||
QY 121 TTTAAGTGCATGTATGTCTGAATCTCATCAAGACAACATGTGTCTGGCTTCAGAAAC 180
DB |||||
QY 244 TTCAAGTGCCACTGCATGTCCGAGTGGCACCAGCGGCAGATGCAGGTGTTCGGCCAGGCC 303
DB |||||
QY 181 CCTCAGCAGTTAATGATTAATTTTTCAGAGGAATTCGGAATGACATTTCTGGAATCTCTG 240
DB |||||
QY 304 CCGACCGAGTGTGCGAGGGCTTCTCGAGGAGTTCCTGACGCCCTTCCTACCTTGCTC 363
DB |||||
QY 241 AGCGAGCGTTTGGCACTAAAGGTCCACAACAACATTTCTTACAAATGAATACATCAGC 300
DB |||||
QY 364 CGCGGCGCCACCGACACTCCCGCATCGCGCCACCGTCTCTACACAGATTCATCGCC 423
DB |||||
QY 301 CACCGAGAGCAGATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAAG 360
DB |||||
QY 424 GACCGCCACCACTCCACATGAATCCACGCGTGGCGCCACGCTACCGAGTTTGTCAAG 483
DB |||||
QY 361 TGCGTGGCAGAGAGGGCTTGTGTAAGGT 390
DB |||||
QY 484 TTCTCGGGCGGAGGGCCACTGCAAGGT 513
DB |||||

RESULT 5

US-10-425-115-157125

;; Sequence 157125, Application US/10425115

;; Publication No. US20040214272A1

;; GENERAL INFORMATION:

;; APPLICANT: La Rosa, Thomas J.

;; APPLICANT: Kovalic, David K.

;; APPLICANT: Zhou, Yihua

;; APPLICANT: Cao, Yongwei

;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

;; TITLE OF INVENTION: Plants

;; FILE REFERENCE: 38-21(53222)B

;; CURRENT APPLICATION NUMBER: US/10/425,115

;; CURRENT FILING DATE: 2003-04-28

;; NUMBER OF SEQ ID NOS: 369326

;; SEQ ID NO 157125

;; LENGTH: 2046

;; TYPE: DNA

;; ORGANISM: Zea mays

;; FEATURE:

;; OTHER INFORMATION: Clone ID: MRT4577_74879C.1

US-10-425-115-157125

Query Match 15.4%; Score 169.2; DB 8; Length 2046;
Best Local Similarity 64.6%; Pred. No. 4.3e-26;
Matches 252; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 1 ATGGCAAGTCGGATTTCTGAGCCCAAGGCCATCGCCAATAGAATTAAAGTCCAAAGGG 60
DB |||||
QY 298 ATGGGGAAGCAGAGTTCCTGACCGCGAAGCGATCGGAATCGGATCAAGGCGAAGGGG 357
DB |||||
QY 61 CTCGAGAAAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCCGCGACGAGATGGC 120
DB |||||
QY 358 CTGCAGAAAGCTGCGGTGGTACTGCCAGATGTGTCAAGAGCAGTGGCCGCGACGAGAACGGG 417
DB |||||
QY 121 TTTAAGTGCATGTATGTCTGAATCTCATCAAGACAACACTGTGTCTGCTTCAGAAAC 180
DB |||||
QY 418 TTCAAGTGCCACTGCATGTCCGAGTGGCACCAGCGGCAGATGCAGGTGTTCGGCATGGCG 477
DB |||||
QY 181 CCTCAGCAGTTAATGATTAATTTTTCAGAGGAATTCGGAATGACATTTCTGGAATCTCTG 240
DB |||||
QY 478 CCGACCGCGTGTGCGAGGGCTTCTCGAGGAATTCCTCGAGTCTTCTCTCCCTCATC 537
DB |||||
QY 241 AGCGAGCGTTTGGCACTAAAGGTCCACAACAACATTTGTCTACAAATGAATACATCAGC 300
DB |||||
QY 538 CGCGCGCGCACCGCCACTCCCGCGTGGCCGCGCACCGTTGTCTACAAAGATACATCGCC 597
DB |||||
QY 301 CACCGAGACACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAAG 360
DB |||||
QY 598 GACCGGCAACCGTCCACATGAACCTCTACGCGTGGGCCACGCTCACCGAGTTCTGTCAG 657
DB |||||
QY 361 TGCGTGGCGAGAGAGGGCTTGTGTAAGGT 390
DB |||||
QY 658 CTCCTGGGCGCGAAGGGTACTGCAAGGT 687
DB |||||

RESULT 6

US-10-425-114-2975

;; Sequence 2975, Application US/10425114

;; Publication No. US20040034888A1

;; GENERAL INFORMATION:

;; APPLICANT: Liu, Jingdong

;; APPLICANT: Zhou, Yihua

;; APPLICANT: Kovalic, David K.

;; APPLICANT: Screen, Steven E.

;; APPLICANT: Tabaska, Jack E.

;; APPLICANT: Cao, Yongwei

;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

;; FILE REFERENCE: 38-21(53113)B

;; CURRENT APPLICATION NUMBER: US/10/425,114

;; CURRENT FILING DATE: 2003-04-28

;; NUMBER OF SEQ ID NOS: 73128

;; SEQ ID NO 2975

;; LENGTH: 1614

;; TYPE: DNA

;; ORGANISM: Zea mays

;; FEATURE:

;; OTHER INFORMATION: Clone ID: 700238628_FLI

US-10-425-114-2975

Query Match 15.3%; Score 168.8; DB 7; Length 1614;
Best Local Similarity 64.7%; Pred. No. 4.7e-26;
Matches 251; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGGGCAAGTCGGATTTCTGAGCCCAAGGCCATCGCCAATAGAATTAAAGTCCAAAGGG 60
DB |||||
QY 80 ATGGGGAAGCAGAGTTCCTGACCGCGAAGCGATCGGAATCGGATCAAGGCGAAGGGG 139
DB |||||
QY 61 CTCGAGAAAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAATGCCGCGACGAGATGGC 120
DB |||||
QY 140 CTGCAGAAAGCTGCGGTGGTACTGTGATGTGTGATGATGTGTGATGATGTGTGATGATG 199
DB |||||
QY 121 TTTAAGTGCATGTATGTCTGAATCTCATCAAGACAACACTGTGTGCTGGCTTCAGAAAC 180
DB |||||
QY 200 TTCAAGTGCCACTGCATGTCCGAGTGGCACCAGGCGAGATGCAGGTGTTCGGCATGGCC 259
DB |||||
QY 181 CCTCAGCAGTTAATGATTAATTTTTCAGAGGAATTCGGAATGACATTTCTGGAATCTCTG 240
DB |||||

260	CCGACCGCGTCGTCGAGGGCTTCTCCGAGAGATTCTCGAGTCCTTCTCTCCCTCATC	319
Db		
241	AGCGCAGCGCTTTTGGCACTAAAAGGCTCCACAACAACATTGTCTCAATGAATACATCAGC	300
Qy		
320	CGCGCGCGCACCGGCACCTCCCGCGTCGCGGCCACCGCTCTACACAGATACATCGCC	379
Db		
301	CACCGGAGACCATCCACATGAACGCTTACCAGCTGGGAGACACTTGACCGACTTTTACCAAG	360
Qy		
380	GACCGGCACCGTCCACATGAATCCACGCGGTGGGCCACGCTCACCGAGTTCGTCGTAAG	439
Db		
361	TGCTGGGCGAGAGAGGGCTTGTGTAAAG	388
Qy		
440	TTCTCGGGCGCGAGGGTACTGTAAAG	467
Db		

```

RESULT 7
US-10-425-114-22614
; Sequence 22614, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 22614
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI
US-10-425-114-22614

```

```

RESULT 8
US-10-739-930-2955
; Sequence 2955, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 2955
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-CLUSTER7583_2
US-10-739-930-2955

```

```

RESULT 9
US-10-424-599-25484
; Sequence 25484, Application US/10424599-
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 25484
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
;

```

OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1
US-10-424-599-25484

Query Match 15.0%; Score 165; DB 7; Length 777;

Best Local Similarity 63.0%; Pred. No. 2.3e-25;
Matches 255; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 1 ATGGGCAAGTCGGATTTCTGAGCCCAAGGCCATCGCCCAATAGATTAAATCCAAAGG 60
DB 140 ATGGGAAATAAGTGTCTTCACACCTAAGCAATTCGCCATCGAATCAAGCAAAAGGA 199
QY 61 CTCACAGAGCTTCGCTGGTACTCCAGATGTGCAAAAGCAATGCCGCGACGAGATGGC 120
DB 200 TTGCAGAGCTTCGGTGGTATTGCCAGATGTGTGAGAGCAGTCCGAGATGAGATGG 259
QY 121 TTTAAGTGCATGTATGTCTGAATCTCATCAAGAACAATGTTGCTGGCTTCAGAAAAC 180
DB 260 TTTAAATGCCATTGCATGAGTGAAGGCCACAGCGTCAATGCGAGATTTTGTGACAAAAC 319
QY 181 CCTCAGCAGTTATGATTTATTTTTCAGAGGATTCGGAATGACATTTCTGGAATCTTG 240
DB 320 CCACACCGGATGTTGAGGCTATTCCGAAGATTTGAGAGTACTTTTCTGGAGCACATG 379
QY 241 AGCGCAGCTTTGGCACTAAAGGGTCCACAACAATTTCTACAAATGAATACATCAGC 300
DB 380 AAGCGAGTACCGATTCACCGCTGTGGCAGCCACTGTAGTTTATACGNATACATAAT 439
QY 301 CACCGAGAGCACATCCACATGAACGCTACCGATGGGAGACACTGACCGACTTTTACCAAG 360
DB 440 GACAGACACCATTCATATGAATCTACTCAGTGGGCTACGCTTACTGAGTTTGTAAAG 499
QY 361 TGCGTGGCAGAGGCTGTGTAAAGGTACCAAGTACAGTGCATGAAG 405
DB 500 TACTTGGGTGCAACTGGCAAAATGTAAGTTGAGGAAACACCCCAAG 544

RESULT 10

US-11-097-143-41039
; Sequence 41039, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41039
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41039

Query Match 14.8%; Score 163; DB 10; Length 3313;
Best Local Similarity 65.0%; Pred. No. 1.2e-24;
Matches 241; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

Query Match 14.8%; Score 163; DB 10; Length 1241;
Best Local Similarity 65.0%; Pred. No. 7.5e-25;
Matches 241; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 1 ATGGGCAAGTCGGATTTCTGAGCCCAAGGCCATCGCCCAATAGATTAAATCCAAAGG 60
DB 69 ATGGGTGCGCGCAGGTAGGTACGCCCAAGTACCTCGCCCAACAAGATGAATCGAAGGGT 128
QY 61 CTCACAGAGCTTCGCTGGTACTCCAGATGTGCAAAAGCAATGCCGCGACGAGATGGC 120
DB 129 TTGCAGAGCTTCGGTGGTATTGCCAGATGTGCGAGAGCAGTCCGCGATGAAACGGC 188
QY 121 TTTAAGTGCATGTATGTCTGAATCTCATCAAGAACAATGTTGCTGGCTTCAGAAAAC 180
DB 189 TTTCAAGTGCACACAGATGAGCGAGTCCCAAGCGCGAGTGTCTCTTTTGGGACAAT 248
QY 181 CCTCAGCAGTTATGATTTATTTTTCAGAGGATTCGGAATGACATTTCTGGAATCTTG 240
DB 249 CTGGCAAAATTCCTGCACAGCTTCAGCAAAAGAGTTCTCCGACGGCTACATGGAGTTGCTG 308
QY 241 AGCGCAGCTTTGGCACTAAAGGGTCCACAACAATTTCTACAAATGAATACATCAGC 300
DB 309 CGCGCGGCTTGGCAGAGGAGCAAGCCAGCGGCCACAGATCTACCAGGATACATTGCC 368
QY 301 CACCGAGAGCACATCCACATGAACGCTACCGATGGGAGACACTGACCGACTTTTACCAAG 360
DB 369 CACAAGGAGCACATCCACATGAAGCCACCGAGTGTCTCACCCTGTCGAGTACGTGAAG 428
QY 361 TGCGTGGGCGAG 371
DB 429 TGGCTGGGCGG 439

RESULT 11

US-11-097-143-41038
; Sequence 41038, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41038
; LENGTH: 3313
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41038

Query Match 14.8%; Score 163; DB 10; Length 3313;
Best Local Similarity 65.0%; Pred. No. 1.2e-24;
Matches 241; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

Db 4006 CTCGAGAGCTGCGCTGGTACTCCAGATGTGCGAGAGCAGTCCGCGATGAACGGC 4065
Qy 121 TTAAAGTGTCACTGTATGTCTGAATCTCATCAAGACAAGTGTGCTGGCTTCAGAAAAC 180
Db 4066 TTCAAGTGTCCACAGATGAGCGAGTCCACAGCGCAGTGTCTCTCTTTGCGGCAAT 4125
Qy 181 CCTCAGAGTTATGATATTTTTCAGAGGAATTCGAAATGACATTTCTGGAATCTTG 240
Db 4126 CCTGGCAAAATTCCTGCACAGCTTCAGCAAAAGATTTCTCCAGCGGTACATGGAGTTGCTG 4185
Qy 241 AGCGAGCTTTGGCACTAAAGGGTCCACAACAAGATTTCTACAAATGAATACATCAGC 300
Db 4186 CGCGGGCGTTGGCAGCAGCAACAGCAGCGCAACAGATCTACAGGAGTACATGGCC 4245
Qy 301 CACGAGAGCAGATCCACATGAAGCTTACCCAGTGGGAGACACTGACCGACTTTACCAAG 360
Db 4246 CACAAGGAGCAGATCCACATGAAGCGCCACCGATGGCTCACCTGTCCGACTACGTGAAG 4305
Qy 361 TGGCTGGGCGAG 371
Db 4306 TGGCTGGGCGG 4316

RESULT 14
US-10-767-701-25587
; Sequence 25587, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 53128
; SEQ ID NO 25587
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30963818
US-10-767-701-25587

Query Match 13.6%; Score 150.4; DB 7; Length 584;
Best Local Similarity 63.6%; Pred. No. 2.8e-22;
Matches 229; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
Qy 1 ATGGGCAAGTCGGAATTTCTGAGCCCAAGGCCATCGCCAATAGAAATTAAGTCCAAAGGG 60
Db 225 ATGGGCAAGTCGGAATTTCTGAGCCCAAGGCCATCGCCAATAGAAATTAAGTCCAAAGGG 284
Qy 61 CTCGAGAGCTTCGCTGTACTGCGAGATGTCCAAAGCAATGCGCGAGAGATGCG 120
Db 285 CTGCAAAAGCTGCGGTGTACTGCGAGATGTCCAAAGCAATGCGCGAGAGATGCG 344
Qy 121 TTAAAGTGTCACTGTATGTCTGAATCTCATCAAGACAAGTGTGCTGGCTTCAGAAAAC 180
Db 345 TTCAAGTGTCACTGTATGTCTGAATCTCATCAAGACAAGTGTGCTGGCTTCAGAAAAC 404
Qy 181 CCTCAGCAGTTTATGGAATTAATTTTTCAGAGGAATTCGAAATGACTTTCTGGAATCTTG 240
Db 405 CCCGACCGGCTGCTGAGGGCTTCTCCGAGGAGTTCTCGAGTCTCTCTCTCCCTCATC 464
Qy 241 AGCGAGCGCTTTGGCACTAAAGGGTCCACAACATTTGCTACAAATGAATACATCAGC 300
Db 465 CGCGCGCGGCAACCGGCACTCCCGCGTCCCGCGCACCGTCTGCTCAACAGAGTACATCGG 524
Qy 301 CACCGAGAGCAGATCCACATGAAGCTACCCAGTGGGAGACACTGACCGACTTTTACCAAG 360
Db 525 GACCGTACCACGTCACATGAATTCACGCGCTGGGCGCACCGCTCACCAGTTCGTCAG 584

RESULT 15
US-09-991-936-1424
; Sequence 1424, Application US/09991936
; Publication No. US20030073827A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
; FILE OF INVENTION: NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-6-C1
; CURRENT APPLICATION NUMBER: US/09/991,936
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US/09/543,668
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,704
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1959
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1424
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-09-991-936-1424
Query Match 11.9%; Score 130.8; DB 3; Length 549;
Best Local Similarity 70.7%; Pred. No. 4.6e-18;
Matches 188; Conservative 0; Mismatches 77; Indels 1; Gaps 1;
Qy 1 ATGGGCAAGTCGGAATTTCTGAGCCCAAGGCCATCGCCAATAGAAATTAAGTCCAAAGGG 60
Db 176 ATGGGTHAAGCAGAGTAGGTACCCCAAGTACATAGCCCAACAAATGAAGGCCAAAGGC 235
Qy 61 CTCGAGAGCTTCGCTGTACTGCGAGATGTCCAAAGCAATGCGCGAGAGATGCG 120
Db 236 CTCGAGAGCTTCGCTGTACTGCGAAATGTGTCAGAAACAATGTCAGAGATGAACCGGT 295
Qy 121 TTAAAGTGTCACTGTATGTCTGAATCTCATCAAGACAAGTGTGCTGGCTTCAGAAAAC 180
Db 296 TTCAATGCCACAAATGTCCGAATCCCAATCCCAATCAAGACAATTTAATCTTTGCTGATAC 355
Qy 181 CCTCAGCAGTTTATGGAATTAATTTTTCAGAGGAATTCGAAATGACTTTCTGGAATCTTG 240
Db 356 TCACACATGTACCTAGATCAATTTTCAAGAATTTCTGCGGCTTCTTAGAATCTTG 414
Qy 241 AGCGAGCGCTTTGGCACTAAAGGGT 266
Db 415 AGAAGACAATTTGGAACCTAAACGAGT 440

Search completed: November 27, 2005, 02:22:26
Job time : 1272.42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:41:08 ; Search time 388.941 Seconds
(without alignments)
421.056 Million cell updates/sec

Title: US-09-555-529-2
Perfect score: 1102
Sequence: 1 atggcgaactcgatttctt.....aaaaaaaaaaaaaaaaaaaa 1102

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3205263 seqs, 74304013 residues

Total number of hits satisfying chosen parameters: 6410526

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA New.*
1: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq2.*
9: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq3.*
10: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48.2	4.4	2120	1	US-10-689-742-163
2	48.2	4.4	2933	1	US-10-131-826A-345
3	47.4	4.3	3501	1	US-10-131-826A-37
4	46.8	4.2	1257	1	US-10-689-742-115
5	46	4.2	1624	1	US-10-131-826A-181
6	45.8	4.2	1108	1	US-10-689-742-209
7	45.6	4.1	1425	1	US-10-689-742-131
8	44.6	4.0	588	1	US-10-689-742-41
9	43.6	4.0	894	1	US-10-689-742-5
10	43.4	3.9	2487	1	US-10-689-742-165
11	43.2	3.9	791	1	US-10-689-742-145
12	43.2	3.9	3819	1	US-10-131-826A-405
13	42.8	3.9	1325	1	US-10-689-742-105
14	42.8	3.9	2033	1	US-10-131-826A-307
15	42.6	3.9	340000	7	US-11-102-978-3
16	42	3.8	1523	1	US-10-131-826A-429
17	41.8	3.8	1658	7	US-11-034-569-15
18	41.8	3.8	2476	1	US-10-131-826A-489
19	41.6	3.8	1904	1	US-10-131-826A-99
20	41.2	3.7	110	1	US-10-689-742-195
21	41.2	3.7	1240	1	US-10-131-826A-91
22	41.2	3.7	4053	1	US-10-131-826A-351
23	41	3.7	2066	1	US-10-523-588-2

ALIGNMENTS

RESULT 1
US-10-689-742-163
; Sequence 163, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: LaVallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 163
; LENGTH: 2120
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-689-742-163
Query Match 4.4%; Score 48.2; DB 1; Length 2120;
Best Local Similarity 51.6%; Pred. No. 0.059;
Matches 110; Conservative 0; Mismatches 103; Indels 0; Gaps 0;
Qy 890 TTTGAAATTTGTATAACAACATTTGAACTGTGAAGCATCAATTTGGTGTAGCCAGG 949
Db 1866 TTTTAAATACCATAATCATGCTATGTAATAATAAGACTACTGGCTTTGTGAGGGAATGT 1925
Qy 950 CATGTGTAACCTACTGTGTAGGGGATTTGTTTGTATTAATAAAAAAATCATCT 1009
Db 1926 TTGTGCAAAATTTTTCCTCTAATGATTAATGTTAAATTTGATTAATAATCTTCAGA 1985
Qy 1010 ATTAAATACCTAGTGAATAGTGGTAAATTTATATAAATCTATCTTTTTTAAAGTG 1069
Db 1986 ATTAAATATCCCTTTTGTCTACTTTTGAACACATAATAATCATTTGTAATCTGCGCTTA 2045
Qy 1070 TAAAAAATAAAAAAATAAAAAAATAAAAAA 1102
Db 2046 AAAAAAATAAAAAAATAAAAAAATAAAAAA 2078

Db 3499 A 3499

RESULT 4

US-10-689-742-115

Sequence 115, Application US/10689742

Publication No. US20050250180A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John M

APPLICANT: Lavallie, Edward R

APPLICANT: Racie, Lisa A

APPLICANT: Evans, Cheryl

APPLICANT: Merberg, David

APPLICANT: Treacy, Maurice

APPLICANT: Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

FILE REFERENCE: 00766.000091.10

CURRENT APPLICATION NUMBER: US/10/689,742

CURRENT FILING DATE: 2003-10-22

PRIOR APPLICATION NUMBER: 09/746,783

PRIOR FILING DATE: 2000-12-21

NUMBER OF SEQ ID NOS: 231

SOFTWARE: Patent in version 3.2

SEQ ID NO 115

LENGTH: 1257

TYPE: DNA

ORGANISM: Homo sapiens

US-10-689-742-115

Query Match 4.2%; Score 46.8; DB 1; Length 1257;

Best Local Similarity 73.2%; Pred. No. 0.097;

Matches 60; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

Qy 1021 AGTGAATAGTGGTAAATTTATATAAAATCATCTATTTTAAAGTGTAAAAA 1080

Db 1172 AGTAGTAATTTAAAGAACCAATAAAACCTCTATTTTATTTTAAATAA 1231

Qy 1081 AAAAAAAAAAAAAAAAAAAAAA 1102

Db 1232 AAAAAAAAAAAAAAAAAAAAAA 1253

RESULT 5

US-10-131-826A-181

Sequence 181, Application US/10131826A

Publication No. US20050245730A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Beresini, Maureen

APPLICANT: DeForge, Laura

APPLICANT: Desnoyers, Luc

APPLICANT: Filvaroff, Ellen

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Sherwood, Steven

APPLICANT: Smith, Victoria

APPLICANT: Stewart, Timothy A.

APPLICANT: Tumas, Daniel

APPLICANT: Watanabe, Colin K

APPLICANT: Wood, William

APPLICANT: Zhang, Zemin

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3330R1C128

CURRENT APPLICATION NUMBER: US/10/131,826A

CURRENT FILING DATE: 2002-04-24

PRIOR APPLICATION NUMBER: 60/049911

PRIOR FILING DATE: 1997-06-18

US-10-131-826A-181

Query Match 4.2%; Score 46; DB 1; Length 1624;

Best Local Similarity 58.1%; Pred. No. 0.16;

Matches 79; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 967 GTGTAGGGGATCTTGTATTATAAAAAAAAAAAATCATCTATTTTAAATAGTGTAA 1086

Db 1469 GTTGATCTTATTTTGTAGCTTGCAAAAGTGACCTTATTTCCAAAGAAATTAATAATGTT 1588

Qy 1027 TAGTTGGGTAAATTTATATAAAATCTATGTTTTTTTAAAGTGTAAAAA 1086

Db 1529 GAATCCAAATCTAGAAATAAAATGAGTTTNNTTCCAAAAA 1588

Qy 1087 AAAAAAAAAAAAAAAAAAAAAA 1102

Db 1589 AAAAAAAAAAAAAAAAAAAAAA 1604

RESULT 6

US-10-689-742-209

Sequence 209, Application US/10689742

Publication No. US20050250180A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John M

APPLICANT: Lavallie, Edward R

APPLICANT: Racie, Lisa A

APPLICANT: Evans, Cheryl

APPLICANT: Merberg, David

APPLICANT: Treacy, Maurice

APPLICANT: Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

FILE REFERENCE: 00766.000091.10

CURRENT APPLICATION NUMBER: US/10/689,742

CURRENT FILING DATE: 2003-10-22

PRIOR APPLICATION NUMBER: 09/746,783

PRIOR FILING DATE: 2000-12-21

NUMBER OF SEQ ID NOS: 231

SOFTWARE: Patent in version 3.2

SEQ ID NO 209

LENGTH: 1108

TYPE: DNA

ORGANISM: Homo sapiens

US-10-689-742-209

```
Query Match 4.2%; Score 45.8; DB 1; Length 1108;
Best Local Similarity 61.2%; Pred. No. 0.15;
Matches 74; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 982 TTTTGTTATTAATAAATAAATCATCTATTAAATACCTAGTGAATAGTTGGTAAATTT 1041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 940 TTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1042 ATAATAAATCTATGTTTTTTTTTAAAGTGTAATAAATAAATAAATAAATAAATAA 1101
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1000 AAGATAGAACATGGTTTGGCTATGCTTGAGAAAAATAAATAAATAAATAAATAA 1059
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1102 A 1102
Db 1060 A 1060

RESULT 7
US-10-689-742-131
; Sequence 131, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: Lavallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; PRIOR FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 131
; LENGTH: 1425
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-689-742-131

Query Match 4.1%; Score 45.6; DB 1; Length 1425;
Best Local Similarity 71.4%; Pred. No. 0.19;
Matches 60; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1019 CTAGTGAATAGTTGGGTAAATTTAATAAATCTATGTTTTTTTAAAGTGTAATAA 1078
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1339 CTGGAACATGGATGGGGATTTGATAGAAAAATAAACCCCTGCTTTTGATTCAAAAA 1398
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1079 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1399 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
US-10-689-742-41
; Sequence 41, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: Lavallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
```

```
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-689-742-41
```

```
Query Match 4.0%; Score 44.6; DB 1; Length 588;
Best Local Similarity 53.1%; Pred. No. 0.22;
Matches 95; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 924 AAGCATCAAAATTTGGTGTAGCCCAAGGCACCTGTAACCTCTACTGTGTAGGGATTTGTT 983
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 AAGAAACCAAGAGGCTAATCACAGGGTTCTGAAATAAATAAATCTGATTTAGATAATTTCTA 415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 984 TTGTATTAAAAAATAAATAAATAAATCATCTATTTTAAATACTAGTGAATAGTTGGGTAAATTTAT 1043
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 416 TAGGAATTTAAAAAATAAATAAATAAATGAAACCTCATGTAAATTTAAGATAGATTCAAACATAAGTCCGA 475
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1044 AATAAAATCTATGTTTTTTTTTAAAGTGTAATAAATAAATAAATAAATAAATAAATAAATAA 1102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 AAGAGGATAGCAAGAATTTTACACAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 9
US-10-689-742-5
; Sequence 5, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: Lavallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 894
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-689-742-5
```

```
Query Match 4.0%; Score 43.6; DB 1; Length 894;
Best Local Similarity 78.8%; Pred. No. 0.42;
Matches 52; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1037 AATTATATAATAAATCTATGTTTTTTTAAAGTGTAATAAATAAATAAATAAATAAATAA 1096
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 AAAATGCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 885
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1097 AAAAAA 1102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 886 AAAAAA 891
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 10
US-10-689-742-165
; Sequence 165, Application US/10689742
; Publication No. US20050250180A1
```

```

Db 752 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 779

RESULT 12
US-10-131-826A-405
; Sequence 405, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 405
; LENGTH: 3819
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-131-826A-405

Query Match 3.9%; Score 43.2; DB 1; Length 3819;
Best Local Similarity 61.6%; Pred. No. 0.88;
Matches 69; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 991 AAAAAAAAAAATCATCTATTTAAATACTAGTGCATAGTTCGGTAAATTTATAATAAAA 1050
Db 3704 AGAAAAATAAAAAATCAGCTGCTTGTAAATCGCTAGCAAAAAAAAAAAAAAAAAAAAA 3763
Qy 1051 TCATATGTTTTTTTAAAGTGTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 1102
Db 3764 AAAAAAAAAAATAAAAAAAAAAATAAAAAAAAAAATAAAAAAAAAAATAAAAAAAAAA 3815

RESULT 13
US-10-689-742-105

```

```

; Sequence 105, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: Lavallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 105
; LENGTH: 1325
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-689-742-105

Query Match          3.9%; Score 42.8; DB 1; Length 1325;
Best Local Similarity 61.6%; Pred. No. 0.73; Indels 1; Gaps 1;
Matches 85; Conservative 0; Mismatches 52;

Qy      966  TGTGTTAGGGGATTGTGTTTGTATTAAAAAATAATCATCTATTAAATACTA-GTG 1024
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      1151  TTTTGTAGTCAATGAACCTTTTGTGCACAAACAGATAAACATCTGTGTTTAACTATGGT 1210
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

Qy      1025  AATAGTTGGGTAAATTTTATAATAAAATCTATGTTTTTTTTTAAGTGTAAAAAATAAAAAA 1084
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      1211  AAACAACATGTTAATGAACATATGCTATCCATGACTTAAATGCACAGTTCAAAAAATAAAA 1270
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

Qy      1085  AAAAAAATAAAAAAATAAAA 1102
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      1271  AAAAAAATAAAAAAATAAAA 1288
          |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 14

US-10-131-826A-307

; Sequence 307, Application US/10131826A

; Publication No. US20050245730A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3330R1C128

; CURRENT APPLICATION NUMBER: US/10/131,826A

; CURRENT FILING DATE: 2002-04-24

; PRIOR APPLICATION NUMBER: 60/049911

; PRIOR FILING DATE: 1997-06-18

; PRIOR APPLICATION NUMBER: 60/056974

; PRIOR FILING DATE: 1997-08-26

```

; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 307
; LENGTH: 2033
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-131-826A-307

Query Match      3.9%; Score 42.8; DB 1; Length 2033;
Best Local Similarity 58.7%; Pred. No. 0.85;
Matches 74; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy      970  TTAGGGGATTTGTTTCTATTAATAAAAAAAAAAATCATCTATTAAATACCTAGTGAATAG 1029
Db      1908  TGAAGCAATATAATATTGTAACAACAAACACAGCTCTTACCTAATAAACATTTTATACT 1967

Qy      1030  TTGGGTAAATTTATAATAAAATCTATGTTTTTTTTTAAAGTGTAAAAAAAAAAAAAAAAAAAAAA 1089
Db      1968  GTTTGTATGTATAAATAAAGGTGCCTTAGCTTTTTCGTAATAAAAAAAAAAAAAAAAAAAAAA 2027

Qy      1090  AAAAAA 1095
Db      2028  AAAAAA 2033

```

```

RESULT 15
US-11-102-978-3
; Sequence 3, Application US/11102978
; Publication No. US20050250142A1
; GENERAL INFORMATION:
; APPLICANT: University of Utah Technology Transfer Office
; APPLICANT: University of Utah Research Foundation
; TITLE OF INVENTION: Diagnosis and Treatment of Herpes Simplex Virus Disease
; FILE REFERENCE: 0274-5537.1US
; CURRENT APPLICATION NUMBER: US/11/102,978
; CURRENT FILING DATE: 2005-04-11
; PRIOR APPLICATION NUMBER: PCT/US2003/033152
; PRIOR FILING DATE: 2003-10-18
; PRIOR APPLICATION NUMBER: 60/419,576
; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 340000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (56948)..(57115)
; OTHER INFORMATION: C21orf34 exon
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (60006)..(81089)
; OTHER INFORMATION: Gene VDACP2; voltage-dependent anion channel isoform 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: (167308)..(167438)

```

; OTHER INFORMATION: C21orf34 exon
; FEATURE:
; NAME/KEY: exon
; LOCATION: (216732)..(216833)
; OTHER INFORMATION: C21orf34 exon
US-11-102-978-3

Query Match 3.9%; Score 42.6; DB 7; Length 340000;
Best Local Similarity 47.9%; Pred. No. 6.3; Indels 0; Gaps 0;
Matches 123; Conservative 0; Mismatches 134;
QY 846 AGTTGAAGGATTCATATGAGACATATCTAAACTTGGCTTGAGTTTGAAAAATTGATAA 905
Db 301737 AGTCAATACATATAGTTTGACTATATATAGTTTATTCATATATAGGATATATGAATAT 301796
QY 906 CAACACATTGAAGCTGTGAAGCATCAAAATGGTGTAGCCAGGCACTGTGTAACCTCTAC 965
Db 301797 ATAAATATAGATATATTTTATTCATATATATCCTATATATAGATATATATAACTATAT 301856
QY 966 TGTGTTAGGGGATTTGTTTGTATTAATAAAAAAATCACTATTTAAATACTAGTGA 1025
Db 301857 GTAAGCCAAATATATATAAAGTAACAATATATTTTACTATATATTTGAATATATAAAT 301916
QY 1026 ATAGTTGGGTAAATTTATATAAAATCTATGTTTTTTTTTAAAGTGTAAAAAATAAAAAA 1085
Db 301917 GTATTTGAATATATATGTTTGAATATATAAATTTATGCAAAATAAAACATATAATAA 301976
QY 1086 AAAAAAAAAAAAAA 1102
Db 301977 ATATATATAGTCAATA 301993

Search completed: November 27, 2005, 02:44:29
Job time : 390.941 secs

This page Blank (used)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:53:47 ; Search time 5325.7 Seconds
(without alignments)
10694.773 Million cell updates/sec

Title: US-09-555-529-3
Perfect score: 1002
Sequence: 1 tgaattcgaagtcgcgtaccgcg.....agcatcaaattggtgttcgc 1002

Scoring table: IDENTITY_NUC
Gapop 10.0 Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters:	11766282
--	----------

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl : ★

ALIGNMENTS

RESULT	ORGAN
AX00331	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORD	
SOURCE	

RESULT 1	LOCUS	1002 bp	DNA	linear	PAT 24-AUG-2000
AX003310	AX003310	Sequence 1	from Patent	W09929845.	

KEYWORDS
SOURCE
Homo sapiens (human)

ORGAN	REFERENCE	AUTHOR	TITLE	JOURNAL
-------	-----------	--------	-------	---------

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

REFERENCE 1
Hominidae; Homo.

REFERENCE
AUTHORS
Angulo-Mora, J.F. and Mauffrey, P.

TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 3 17-JUN-1999;

ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)

FEATURES	Location/Qualifiers
1. <i>the</i>	1. <i>the</i>
2. <i>the</i>	2. <i>the</i>
3. <i>the</i>	3. <i>the</i>
4. <i>the</i>	4. <i>the</i>
5. <i>the</i>	5. <i>the</i>
6. <i>the</i>	6. <i>the</i>
7. <i>the</i>	7. <i>the</i>
8. <i>the</i>	8. <i>the</i>
9. <i>the</i>	9. <i>the</i>
10. <i>the</i>	10. <i>the</i>
11. <i>the</i>	11. <i>the</i>
12. <i>the</i>	12. <i>the</i>
13. <i>the</i>	13. <i>the</i>
14. <i>the</i>	14. <i>the</i>
15. <i>the</i>	15. <i>the</i>
16. <i>the</i>	16. <i>the</i>
17. <i>the</i>	17. <i>the</i>
18. <i>the</i>	18. <i>the</i>
19. <i>the</i>	19. <i>the</i>
20. <i>the</i>	20. <i>the</i>
21. <i>the</i>	21. <i>the</i>
22. <i>the</i>	22. <i>the</i>
23. <i>the</i>	23. <i>the</i>
24. <i>the</i>	24. <i>the</i>
25. <i>the</i>	25. <i>the</i>
26. <i>the</i>	26. <i>the</i>
27. <i>the</i>	27. <i>the</i>
28. <i>the</i>	28. <i>the</i>
29. <i>the</i>	29. <i>the</i>
30. <i>the</i>	30. <i>the</i>
31. <i>the</i>	31. <i>the</i>
32. <i>the</i>	32. <i>the</i>
33. <i>the</i>	33. <i>the</i>
34. <i>the</i>	34. <i>the</i>
35. <i>the</i>	35. <i>the</i>
36. <i>the</i>	36. <i>the</i>
37. <i>the</i>	37. <i>the</i>
38. <i>the</i>	38. <i>the</i>
39. <i>the</i>	39. <i>the</i>
40. <i>the</i>	40. <i>the</i>
41. <i>the</i>	41. <i>the</i>
42. <i>the</i>	42. <i>the</i>
43. <i>the</i>	43. <i>the</i>
44. <i>the</i>	44. <i>the</i>
45. <i>the</i>	45. <i>the</i>
46. <i>the</i>	46. <i>the</i>
47. <i>the</i>	47. <i>the</i>
48. <i>the</i>	48. <i>the</i>
49. <i>the</i>	49. <i>the</i>
50. <i>the</i>	50. <i>the</i>
51. <i>the</i>	51. <i>the</i>
52. <i>the</i>	52. <i>the</i>
53. <i>the</i>	53. <i>the</i>
54. <i>the</i>	54. <i>the</i>
55. <i>the</i>	55. <i>the</i>
56. <i>the</i>	56. <i>the</i>
57. <i>the</i>	57. <i>the</i>
58. <i>the</i>	58. <i>the</i>
59. <i>the</i>	59. <i>the</i>
60. <i>the</i>	60. <i>the</i>
61. <i>the</i>	61. <i>the</i>
62. <i>the</i>	62. <i>the</i>
63. <i>the</i>	63. <i>the</i>
64. <i>the</i>	64. <i>the</i>
65. <i>the</i>	65. <i>the</i>
66. <i>the</i>	66. <i>the</i>
67. <i>the</i>	67. <i>the</i>
68. <i>the</i>	68. <i>the</i>
69. <i>the</i>	69. <i>the</i>
70. <i>the</i>	70. <i>the</i>
71. <i>the</i>	71. <i>the</i>
72. <i>the</i>	72. <i>the</i>
73. <i>the</i>	73. <i>the</i>
74. <i>the</i>	74. <i>the</i>
75. <i>the</i>	75. <i>the</i>
76. <i>the</i>	76. <i>the</i>
77. <i>the</i>	77. <i>the</i>
78. <i>the</i>	78. <i>the</i>
79. <i>the</i>	79. <i>the</i>
80. <i>the</i>	80. <i>the</i>
81. <i>the</i>	81. <i>the</i>
82. <i>the</i>	82. <i>the</i>
83. <i>the</i>	83. <i>the</i>
84. <i>the</i>	84. <i>the</i>
85. <i>the</i>	85. <i>the</i>
86. <i>the</i>	86. <i>the</i>
87. <i>the</i>	87. <i>the</i>
88. <i>the</i>	88. <i>the</i>
89. <i>the</i>	89. <i>the</i>
90. <i>the</i>	90. <i>the</i>
91. <i>the</i>	91. <i>the</i>
92. <i>the</i>	92. <i>the</i>
93. <i>the</i>	93. <i>the</i>
94. <i>the</i>	94. <i>the</i>
95. <i>the</i>	95. <i>the</i>
96. <i>the</i>	96. <i>the</i>
97. <i>the</i>	97. <i>the</i>
98. <i>the</i>	98. <i>the</i>
99. <i>the</i>	99. <i>the</i>
100. <i>the</i>	100. <i>the</i>

```
source
1..1002
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
```

ORIGIN

Query Match 100.0%; Score 1002; DB 6; Length 1002;

Best Local Similarity	100.0%	Pred. No. 2.9e-227;	
Matches 1002; Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;

1 TGATTCGAGCTCGGTACCCGGGGATCCGATTAGAAAGTGATCGTGGGTGGTGGCCATG 60

1 TGATTGAGTCGGTACCCGGGGATCCGATTAGAAAGTATCGCTGCCGTGGTCGCATG 60

61 GGGAAGTCGATTTTCTTACTCCCAAGGCTATCGCCAACAGGATCAAGTCCAAGGGCTG 120

Db 61 GGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAGGGGGCTG 120

Qy 121 CAGAAGCTACGCTGGTATTGCCAGATGTGCCAGAGCACTGCCGGGACAGAAATGGCTTT 180

Db
121 CAGAAGCTACGTTGGTATTGCCAGATGTGCCAGAACAGTCCCGGGACGAGATGGCTTT 180

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1002	100.0	1002	6	AX003310	Sequence
2	704.6	70.3	1102	6	AX003309	Sequence
3	653.4	65.2	1528	8	AX003308	Sequence
4	653.4	65.2	1528	8	HBAJ5273	Homo sapi
5	638.4	63.7	2069	8	BC017309	Homo sapi
6	470	46.9	1002	6	AX003311	Sequence
7	438	43.7	1117	6	CQ722054	Sequence
8	425.8	42.5	1414	9	MMKIN17	Mouse KIN17
9	422.4	42.2	1386	9	BC028860	Mus muscu
10	422.4	42.2	221547	14	AC119716	Rattus no
11	416.6	41.6	1390	6	AX003331	Sequence
12	390.6	39.0	2710	9	BC058169	Mus muscu
13	390.6	39.0	180882	14	AC114615	Mus muscu
14	390.6	39.0	182400	14	AC124010	Mus muscu
15	390.6	39.0	185378	9	AL772367	Mouse DNA
16	384.4	38.4	144118	14	AC141995	Rattus no
17	371.2	37.0	.461	6	CQ684251	Sequence
18	354	35.3	130262	14	AC166154	Mus muscu

QY 181 AAGTGCTATTGTATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTCAGAAATCCT 240
Db |||||
QY 181 AAGTGCTATTGTATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTCAGAAATCCT 240
Db |||||
QY 241 CAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTCAGG 300
Db |||||
QY 241 CAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTCAGG 300
Db |||||
QY 301 AGACGCTTTGGCACTAAAGGGTCCAAACAATTTGTCTCAACGAATACATCAGCCAC 360
Db |||||
QY 301 AGACGCTTTGGCACTAAAGGGTCCAAACAATTTGTCTCAACGAATACATCAGCCAC 360
Db |||||
QY 361 CGAGAGCACAATCCACATGAATGCCACTAGTGGGAATCTGACTGATTTTACTAAGTG 420
Db |||||
QY 421 CTGGGCGAGAGAAGGCTTTGTGCAAAAGTGCATGAGACGATAGGAAGTTTCAGCATCAGTG 480
Db |||||
QY 421 CTGGGCGAGAGAAGGCTTTGTGCAAAAGTGCATGAGACGATAGGAAGTTTCAGCATCAGTG 480
Db |||||
QY 481 AAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAGAAATCT 540
Db |||||
QY 481 AAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAGAAATCT 540
Db |||||
QY 541 GCACCTGGATGAAATCATGAGATTTGAAGAGGAAAGAAAGAACTGCCGACAGACTAC 600
Db |||||
QY 541 GCACCTGGATGAAATCATGAGATTTGAAGAGGAAAGAAAGAACTGCCGACAGACTAC 600
Db |||||
QY 601 TGGCTACAGCTGAAATTTATGTAATTTAAATGAGGCTACAGAGGAATGAAAGTACCTAGAAATCC 840
Db |||||
QY 601 TGGCTACAGCTGAAATTTATGTAATTTAAATGAGGCTACAGAGGAATGAAAGTACCTAGAAATCC 840
Db |||||
QY 841 ATCAATGAGAAGACTTTTTCAGTACTATCGTCAATGAAACTGGCCCTTTAAAGAGAGCG 900
Db |||||
QY 841 ATCAATGAGAAGACTTTTTCAGTACTATCGTCAATGAAACTGGCCCTTTAAAGAGAGCG 900
Db |||||
QY 901 AGAGTTGAAGGAATTCATATGAGACATTTCTAACTTGCCTGAGTTTGAATTTGTT 960
Db |||||
QY 901 AGAGTTGAAGGAATTCATATGAGACATTTCTAACTTGCCTGAGTTTGAATTTGTT 960
Db |||||
QY 961 AACAAATACCTTTAAATCTTAAAGCATCAATTTGGTTCGC 1002
Db |||||
QY 961 AACAAATACCTTTAAATCTTAAAGCATCAATTTGGTTCGC 1002
Db |||||

RESULT 2
AX003309
LOCUS AX003309 1102 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 2 from Patent WO9929845.
ACCESSION AX003309
VERSION AX003309.1 GI:9927126
KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1
REFERENCE
AUTHORS Angulo-Mora, J.F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 2 17-JUN-1999;

FEATURES
source location/Qualifiers
1..1102
/organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"
ORIGIN
Query Match 70.3%; Score 704.6; DB 6; Length 1102;
Best Local Similarity 86.1%; Pred. No. 1.4e-156;
Matches 819; Conservative 0; Mismatches 119; Indels 13; Gaps 3;
QY 58 ATGGGGAAGTCCGATTTTCTTACTCCCAAGGCTATCCGCAACAGGATCAAGTCCCAAGGG 117
Db 1 ATGGGCAAGTCCGATTTTCTGAGCCCCCAAGGCCATAGAAATTAAGTCCCAAGGG 60
QY 118 CTGCAGAAAGCTACGCTGGTATTGGCCAGATGTGCCAAGAGCAGTGCCTGGGACGAGAAATGCG 177
Db 61 CTCGAAAGCTTGGCTGGTACTGCGAGATGTGCCAAGAGCAATGCCGCGACGAGAAATGCG 120
QY 178 TTTAAGTGTCAATGTGATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTCAGAAAT 237
Db 121 TTTAAGTGTCAATGTGATGTCGGAATCTCATCAAGAGCAACTGTTGCTGGCTTCAGAAAT 180
QY 238 CCTCAGAGATTTATGGATTTATTTTTCAGAGAAATTCGGAATGACTTTCTAGAACTTCTC 297
Db 181 CCTCAGAGATTTATGGATTTATTTTTCAGAGAAATTCGGAATGACTTTCTGGAATCTTCTG 240
QY 298 AGGAGACCTTTGGCAGCTAAAGGGTCCCAACAACATTTGTCTACAACGAATACATCAGC 357
Db 241 AGGAGACCTTTGGCAGCTAAAGGGTCCCAACAACATTTGTCTACAATGAAATACATCAGC 300
QY 358 CACGAGAGCAGCATCCACATGAATGCCACTCAGTGGGAACTCTGAGCTGATTTTACTAAG 417
Db 301 CACGAGAGCAGCATCCACATGAACGCTACCCAGTGGGAGACACTGACCGCACTTTACCAAG 360
QY 418 TGGCTGGGCGAGAGAGGCTTGTGCAAA-----AGTCACTGAGAGACATAGGAAGTTCA 471
Db 361 TGGCTGGGCGAGAGAGGCTTGTGTAAGGTTACCAAGTGCATGGAAGCTCTGCGGAGCGCA 420
QY 472 GCATCAGTGAACGAGAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAGAGAA 531
Db 421 GCATCCGGGAAACGGAAGAGCTTTCAGAGACTTCGCGCCAGCCT-----GCGAAGAG 474
QY 532 AAGAAATCTGCACTGGATGAAATCATGGAGATTGAAGAGGAAAGAAAGAACTCCCGA 591
Db 475 AAGAAATCGGCGCTGGATGAGATCATGGAGCTCGAAGAGGAAAGAAAGAACCCACGG 534
QY 592 ACAGACTACTGGCTACAGCTGAAATTTATTTGTAAGAAATTAACCAAGAAACTGGGAG 651
Db 535 ACAGACGCTTGTGTTACAGCCGGGATCGTTGTGAAATTTAACGAAAGCTTGGGAG 594
QY 652 AAATATCATAGAAAGGCTATTGTTAAGGAAGTAAATTGAAATATATACAGCTCTGTG 711
Db 595 AAATATCATAGAAAGGCTCGTTAAGGAAGTGAATTGACAGGTACACAGCTGTGTA 654
QY 712 AAGATGATTGATTTCTGGAGACAAAGCTGAAACTTTCAGCTACTACTCGTCATTTTAGACAGTAAT 771
Db 655 AAGATGATGATCTCTGGAGACAGGCTGAAACTTGGACACAGACTCAATTTAGACAGACTATT 714
QY 772 CCAGCACGAGGAAAGAAATTTCTAGTTTTTAATGGAGGCTACAGAGGAATGAGGTACC 831
Db 715 CCGGCCCGGGGAAAGGGTTCTAGTTTTTAATGGAGGCTACAGAGGAATGAGGCACT 774
QY 832 CTAGAAATCCATCAATGAGAACTTTTTTTCAGCTACTACTCGTCATTTGAAACTGGCCCTTTA 891
Db 775 CTCGAATCCATCAATGAGAGGCTTTTTTCAGCCAGGATAGTCAATTTGAAACTGGACTTTG 834
QY 892 AAAGGACGAGAGTTGAAGGAATTCATATGAAGACATTTCTTAACTTGCCTGAGTTGA 951
Db 835 AAAGGACGAGAGTTGAAGGATTTCAATATGAAGACATATCTTAACTTGCCTGAGTTGA 894


```

Qy 952 AAATTTGTTAAGCAATACCTTTAAATCTTAAAGCATCAAAATTTGTTGTCG 1002
Db 895 AAATTTGTTAAGCAATACCTTTAAATCTTAAAGCATCAAAATTTGTTGTTAGC 944

RESULT 3
AX003308
LOCUS AX003308 1528 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 1 from Patent WO929845.
ACCESSION AX003308
VERSION AX003308.1 GI:9927125
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Angulo-Mora, J.F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 929845-A 1 17-JUN-1999;
(FR) ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
FEATURES
source Location/Qualifiers
1..1528
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 65.2%; Score 653.4; DB 6; Length 1528;
Best Local Similarity 76.6%; Pred. No. 2e-144;
Matches 969; Conservative 0; Mismatches 1; Indels 295; Gaps 2;

Qy 32 AGAAAGTGATCGCTGCGCGTGTGCGCATGGGGAAGTCGGAATTTCTTACTCCCAAGGCTA 91
Db 40 AGAAAGTGATCGCTGCGCGTGTGCGCATGGGGAAGTCGGAATTTCTTACTCCCAAGGCTA 99

Qy 92 TCGCCAAACAGATCAAGTCCAAAGGGGTGCGAAGCTACGCTGGTATTCAGAGATGTGCC 151
Db 100 TCGCCAAACAGATCAAGTCCAAAGGGGTGCGAAGCTACGCTGGTATTCAGAGATGTGCC 159

Qy 152 AGAAGCAGTCCGGGAGAGAGATGGCTTTAAGTGTCTATGTTATGTCGGAATCTCATAGA 211
Db 160 AGAAGCAGTCCGGGAGAGAGATGGCTTTAAGTGTCTATGTTATGTCGGAATCTCATAGA 219

Qy 212 GACAACTATTGCTGCTTCAGAAATCCTCAGCAGTTCATGGAATTTTTCAGAGGAAT 271
Db 220 GACAACTATTGCTGCTTCAGAAATCCTCAGCAGTTCATGGAATTTTTCAGAGGAAT 279

Qy 272 TCCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAAGGGTCCACAACA 331
Db 280 TCCGAAATGACTTTCTAGAACTTCTCAGGAGACGCTTTGGCACTAAAGGGTCCACAACA 339

Qy 332 ACATTGTCTACAGAAATACATCAGCCAGGAGCAGATCCATCAATGAATGCCACTCAGT 391
Db 340 ACATTGTCTACAGAAATACATCAGCCAGGAGCAGATCCATCAATGAATGCCACTCAGT 399

Qy 392 GGGAAACTCTGACTGATTTTACTAAGTGGCTGGGAGAGAGGGCTTTGTGCAAA----- 444
Db 400 GGGAAACTCTGACTGATTTTACTAAGTGGCTGGGAGAGAGGGCTTTGTGCAAAAGTGGACG 459

Qy 445 ----- 444
Db 460 AGACACAAAGGCTGGTATATTTCAGTACATAGACAGGAGCCAGAAACTATCCGCGGC 519

Qy 445 ----- 444
Db 520 AACTGGAACTGGAGAAAAAGAAAAGCAGGACCTTTGATGATGAAGAAAAAAATGCGCAAAAT 579

Qy 445 ----- 444
Db 580 TTATTGAAGCAAGTGAAGAGAGCGCTTGGAGGGAAGCAAGAGAGGTCCTTACTTTTA 639

```

```

Qy 445 ----- 444
Db 640 CGGAATTAAGCAGAGAAAATGATGAAGAGAAAGTCACGTTTAATTTGAGTAAAGGAGCAT 699

Qy 445 ----- 444
Db 700 GTAGCTCATCCGGAGCAACATCTTCCAAGTCAAGTACTCTCTGGGACCGAGTGCACTGAAGA 759

Qy 458 CGATAGGAGTTTCAGCATCAGTCAGAACAGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTA 517
Db 760 CGATAGGAGTTTCAGCATCAGTCAGAACAGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTA 819

Qy 518 AAGAAAGAGAGAAAAGAAATCTGCACTCGATGAAATCATGGAGATTTGAAGAGAGAAAGA 577
Db 820 AAGAAAGAGAGAAAAGAAATCTGCACTCGATGAAATCATGGAGATTTGAAGAGAGAAAGA 879

Qy 578 AAAGAACTGCCGAAACAGACTACTGGCTACAGCTGAGCTGAAATTTATTTGAAAAATTAACCA 637
Db 880 AAAGAACTGCCGAAACAGACTACTGGCTACAGCTGAGCTGAAATTTATTTGAAAAATTAACCA 939

Qy 638 AGAACTGGGAGAGAAATATCATAGAAAGAGGCTATTTGTTAAGGAGATTAATTGACAAAT 697
Db 940 AGAACTGGGAGAGAAATATCATAGAAAGAGGCTATTTGTTAAGGAGATTAATTGACAAAT 999

Qy 698 ATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAGCTGAAACTTTGACAGACTCAT 757
Db 1000 ATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAGCTGAAACTTTGACAGACTCAT 1059

Qy 758 TAGAGACAGTAATTTCCAGCACAGGAGAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAG 817
Db 1060 TAGAGACAGTAATTTCCAGCACAGGAGAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAG 1119

Qy 818 GAAATGAGGTACCTTAGAATCCATCAATGAGAGAGCTTTTTCAGCTACTATCGTCAATG 877
Db 1120 GAAATGAGGTACCTTAGAATCCATCAATGAGAGAGCTTTTTCAGCTACTATCGTCAATG 1179

Qy 878 AAATGCGCCCTTTAAAGGAGCAGAGTTCGAAGGAATTCATATGAAGACATTTTCTAAAC 937
Db 1180 AAATGCGCCCTTTAAAGGAGCAGAGTTCGAAGGAATTCATATGAAGACATTTTCTAAAC 1239

Qy 938 TTGCTGAGTTGAAAATTTGTTAAACAATACCTTTTAAATCTTTAAAGCATCAAAATTTGGTG 997
Db 1240 TTGCTGAGTTGAAAATTTGTTAAACAATACCTTTTAAATCTTTAAAGCATCAAAATTTGGTG 1298

Qy 998 TTGCGC 1002
Db 1299 TTGCGC 1303

RESULT 4
HSAJ5273
LOCUS HSAJ5273 1528 bp mRNA linear PRI 13-DEC-2000
DEFINITION Homo sapiens mRNA for Kin17 protein.
ACCESSION AJ005273
VERSION AJ005273.1 GI:3850703
KEYWORDS KIN17 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Kannouche, P., Mauffrey, P., Pinon-Lataillade, G., Mattei, M.G.,
Sarasin, A., Daya-Grosjean, L. and Angulo, J.F.
TITLE Molecular cloning and characterization of the human KIN17 cDNA
encoding a component of the UVC response that is conserved among
metazoans
JOURNAL Carcinogenesis 21 (9), 1701-1710 (2000)
PUBMED 10964102
REFERENCE 2
AUTHORS Mauffrey, P.
TITLE Direct Submission

```

JOURNAL	Submitted (08-APR-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 - Fontenay aux Roses Cedex, 92265, FRANCE
REMARK	Revised by [3]
REFERENCE	3 (bases 1 to 1528)
AUTHORS	Mauffrey, P.
TITLE	Direct Submission
JOURNAL	Submitted (07-OCT-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 - Fontenay aux Roses Cedex, 92265, FRANCE
COMMENT	On Nov 7, 1998 this sequence version replaced gi:3046739. Kannouche P., Mauffrey P., Pinon-Lataillade G., Biard D.S.F., Angulo J.F.
FEATURES	Sequences codant pour la Proteine kin17 et leurs Applications - French Patent Nr 97 15536 1:1-50(1997).
	Location/Qualifiers
	1..1528
	/organism="Homo sapiens"
	/mol_type="mRNA"
gene	/db_xref="taxon:9606"
	/sex="male"
	/tissue_type="testis"
	1..1528
	/gene="Kin17"
CDS	66..1247
	/gene="Kin17"
	/codon_start=1
	/protein_id="CAA06462.1"
	/db_xref="GI:3850704"
ORIGIN	/db_xref="GOA:O60870"
	/db_xref="UniProt/TREMBL:O60870"
	/translation="MGKSDPLPKAIFANIKSKGLQKLWYCOMKOCRDENGPKCH CMSEHQROLLLASNPQDFMDFSEEPFNDFLELLRRFGYKRVHNNIVNEYISHR EHHNMATWETLTDFTKWLREGLECKDETPKMWIYQIDRPETIRKQLELEKKKK QLDDBEKTAKTIEEQVRRGLEKEQEVPTFELRENDDEERVTNLSKGCSSGAT SSKSTLGSALKITGSSASVKRKSQSSQSTQKSKKKKKSALDEIMETEEBKRTAR TDYWLQPEIIVKIITKLGKVKHKKAIKVEVIDKYTAVVKIMIDSGDKLKLQDTHLET VTPAPGKRILVLNGGYRNGNEGLESINEKTFSAIVIEIETGPKLGRVVGIOVEDISKL A"
	Query Match 65.2%; Score 653.4; DB 8; Length 1528;
	Best Local Similarity 76.6%; Pred. No. 2e-144;
QY	Matches 969; Conservative 0; Mismatches 1; Indels 295; Gaps 2;
	32 AGAAAGTGATCGTCGCGTGGTCCGATCGGGAAGTCGGAATTTCTTATCCCAAGGCTA 91
	40 AGAAAGTGATCGTCGCGTGGTCCGATCGGGAAGTCGGAATTTCTTATCCCAAGGCTA 99
QY	92 TCGCCAAACAGGATCAAGTCCAAAGGGCTGCAGAGAGCTACGCTGGTATTCGCCAGATGCGC 151
Db	
QY	100 TCGCCAAACAGGATCAAGTCCAAAGGGCTGCAGAGAGCTACGCTGGTATTCGCCAGATGCGC 159
Db	
QY	152 AGAAGCAGTGCCTGGGACGAGATGGCTTTAAAGTGTCATTTATGTCGCGAATCTCATCAGA 211
Db	
QY	160 AGAAGCAGTGCCTGGGACGAGATGGCTTTAAAGTGTCATTTATGTCGCGAATCTCATCAGA 219
Db	
QY	212 GACAACTATGTGCGCTTCAGAAAAATCTCAGCAGGTTATGGAATATTTTCAGAGGAAT 271
Db	
QY	220 GACAACTATGTGCGCTTCAGAAAAATCTCAGCAGGTTATGGAATATTTTCAGAGGAAT 279
Db	
QY	272 TCCGAATAGCTTTCTAGAACTTCTCAGGAGAGCTTTGGCACTAAAGGGTCCACAACA 331
Db	
QY	280 TCCGAATAGCTTTCTAGAACTTCTCAGGAGAGCTTTGGCACTAAAGGGTCCACAACA 339
Db	
QY	332 ACATTGCTACACGAATACATCAGCCACCGAGAGCAGATCCACATGAATGCCACTCAGT 391
Db	
QY	340 ACATTGCTACACGAATACATCAGCCACCGAGAGCAGATCCACATGAATGCCACTCAGT 399
Db	
QY	392 GGGAAACTCTGACTGATTTTACTTAAGTGGCTGGGACAGAGAGCTTGTGCAAA----- 444
Db	
QY	400 GGGAAACTCTGACTGATTTTACTTAAGTGGCTGGGACAGAGAGCTTGTGCAAAAGTGAGCG 459
Db	
QY	445 ----- 444

Db	460 AGACACCAAAAGGCTGGTATATTTCAGTACATAGACAGGGACCCAGAAAACATATCCCGCGC 519
QY	445 ----- 444
Db	520 AACTGGAACCTGGAGAAAAAGAAAGCAGGACCTTGTATGATGATGAAGAAAAAACTGCCAAAT 579
QY	445 ----- 444
Db	580 TTATTGAAGACCACTGAGAGAGGCGCTGGAGGGAAGAAACAGGAGGTCCCTACTTTTA 639
QY	445 ----- 444
Db	640 CGGAATTAAAGCAGAAAAATGATGAAGAGAAAGTCACGTTTAAATTTTGATGAAGAGCAT 699
QY	445 -----AGTGCACTGAAGA 457
Db	700 GTAGCTCATCCGGAGCAACATCTTCCAAGTCAAGTACTCTGGGACCGAGTGCACTGAAGA 759
QY	458 CGATAGGAAGTTTCAGCATCAGTGAACGAAAAAGAAATCTTCCAGAGCTCAACTAGTCTTA 517
Db	760 CGATAGGAAGTTTCAGCATCAGTGAACGAAAAAGAAATCTTCCAGAGCTCAACTAGTCTTA 819
QY	518 AAGAAAAGAGAAAAAGAAATCTGCACCTGGATGAAATCATGGAGATTGAAGAGGAAAAAGA 577
Db	820 AAGAAAAGAGAAAAAGAAATCTGCACCTGGATGAAATCATGGAGATTGAAGAGGAAAAAGA 879
QY	578 AAAGAACTGCCCGAACACAGACTACTGGCTACAGCCTGAAATTTATTTGAAAAATTATAACCA 637
Db	880 AAAGAACTGCCCGAACACAGACTACTGGCTACAGCCTGAAATTTATTTGAAAAATTATAACCA 939
QY	638 AGAAACTGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAAGTAAATTTGACAAAT 697
Db	940 AGAACTGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAAGTAAATTTGACAAAT 999
QY	698 ATACAGCTGTTGTTGAAGATGATTGATTCTCGAGACAACTGAAATCTGACCAAGACTCAT 757
Db	1000 ATACAGCTGTTGTTGAAGATGATTGATTCTCGAGACAACTGAAATCTGACCAAGACTCAT 1059
QY	758 TAGACACAGTAATTCACGACCCAGGAAAAAGAAATTTCTAGTTTTTAAATGGAGCTACAGAG 817
Db	1060 TAGACACAGTAATTCACGACCCAGGAAAAAGAAATTTCTAGTTTTTAAATGGAGCTACAGAG 1119
QY	818 GAAATGAAGGTACCTTAGAATCCATCAATGAGAACACTTTTTCAGCTACTATCGTCATTG 877
Db	1120 GAAATGAAGGTACCTTAGAATCCATCAATGAGAACACTTTTTCAGCTACTATCGTCATTG 1179
QY	878 AAACCTGGCCCTTTAAAAGGACGACAGATTTGAAGGAATTCATATGAAGACATTTTCTAAAC 937
Db	1180 AAACCTGGCCCTTTAAAAGGACGACAGATTTGAAGGAATTCATATGAAGACATTTTCTAAAC 1239
QY	938 TTGCTGAGTTTGAATAATTTGTTAAACAATACCTTTTAAATCTTAAAGCATCAAAATGGTG 997
Db	1240 TTGCTGAGTTTGAATAATTTGTTAAACAATA-CATTTAAATCTTAAAGCATCAAAATGGTG 1298
QY	998 TTCGC 1002
Db	1299 TTCGC 1303
RESULT 5	
BC017309	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
1 (bases 1 to 2069)	
BC017309	
Homo sapiens	
KIN, antigenic determinant of recA protein homolog	
(mouse), mRNA (cdna clone MGC:29595 IMAGE:5089578), complete cds.	
BC017309	
Homo sapiens	
MGC.	
Homo sapiens (human)	
Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;	
Hominidae; Homo.	
1 (bases 1 to 2069)	

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L.H., Shemen, C.M., Schuler, G.D., Altschul, S.F., Zebberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Rahay, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E., Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 2069)

Strausberg, R.

Direct Submission

Submitted (05-NOV-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

On Dec 19, 2003 this sequence version replaced gi:16878214.

Contact: MGC help desk

Email: cgabs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@nhgri.nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,

Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,

Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,

Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,

Maduro, Q.L., Masiello, C., Maekeri, B., Mastrian, S.D., McCloskey, J.C.,

McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,

Tourgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,

Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 39 Row: O Column: 18

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 13124882.

FEATURES

source

1. :2069
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:29595 IMAGE:5089578"
/tissue_type="pancreas, epithelioid carcinoma"
/clone_lib="NIH MGC 42"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

gene

1. :2069
/gene="KIN"
/note="synonyms: KIN17, BTCD"
/db_xref="GeneID:22944"
/db_xref="MIM:601720"

CDS

12. :1193
/gene="KIN"
/codon_start=1

ORIGIN

Query Match	63.7%	Score 638.4	DB 8	Length 2069
Best Local Similarity	76.3%	Pred. No. 7.3e-141		
Matches 954	Conservative 0	Mismatches 1	Indels 295	Gaps 2

```

47  CCGTGGTCCCATGGGAAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCA 106
      |||||
1  CCGTGGTCCCATGGGAAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCA 60
      |||||

107 AGTCCAAAGGGGCTGCAGAAAGCTACGCTGTATTCAGATGTCAGAGAGAGAGAGAGAGAG 166
      |||||
61 AGTCCAAAGGGGCTGCAGAAAGCTACGCTGTATTCAGATGTCAGAGAGAGAGAGAGAGAG 120
      |||||

167 ACAGAAATGGCTTTAAGTGTCAATGTCGGAATCTCATCAGAGACAACTATTGCTGG 226
      |||||
121 ACAGAAATGGCTTTAAGTGTCAATGTCGGAATCTCATCAGAGACAACTATTGCTGG 180
      |||||

227 CTTCAGAAATCCTCAGCAGCTTTATGGATTTATTTTTCAGAGGAAATCCGAAATGACTTTC 286
      |||||
181 CTTCAGAAATCCTCAGCAGCTTTATGGATTTATTTTTCAGAGGAAATCCGAAATGACTTTC 240
      |||||

287 TAGAATCTTTCAGAGAGACGCTTTGGCACTAAAGAGGTCCAAACAACTTGTCTACACG 346
      |||||
241 TAGAATCTTTCAGAGAGACGCTTTGGCACTAAAGAGGTCCAAACAACTTGTCTACACG 300
      |||||

347 AATACATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAACTCTGACTG 406
      |||||
301 AATACATCAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAACTCTGACTG 360
      |||||

407 ATTTTACTAAGTGGCTGGGAGAGAGGCTTTGTGCAAA 444
      |||||
361 ATTTTACTAAGTGGCTGGGAGAGAGGCTTTGTGCAAAAGTGGAGACACCAAAAGGCT 420
      |||||
445 ----- 444
      |||||

421 GGTATATTCACTACATAGACAGGAGCCAGAAACTATCCCGCGGCACTGGAACTGGAGA 480
      |||||
445 ----- 444
      |||||

481 AAAAGAAAAGCAGGACCTTTGATGATGAAGAAAACCTGCAAAATTTTATGAAGAGCAAG 540
      |||||
445 ----- 444
      |||||

541 TGAGAGAGGCGCTGGAAGGAGGAGGAGGAGGCTCTCTTTTACGGAATTAAGCAGAG 600
      |||||
445 ----- 444
      |||||

601 AAAATGATGAAGAGAAAGTCAACGTTTAAATTTGAGTAAGGAGGATGTAGCTCATCCGAG 660
      |||||
445 ----- 444
      |||||

661 CAACATCTTCCAAAGTCAAGTACTCTCTGGGACCGAGTGCATGAGACCATAGGAGTTTCA 720
      |||||
473 CATCAGTGAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAAGAGAAA 532
      |||||
721 CATCAGTGAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAAGAGAAA 780
      |||||
533 AGAAATCTGCACTGGATGAAATCATGGAGATTGAAGAGGAGAAAAGAAAGAACTCCCGAA 592
      |||||

```

Db 781 AGAATCTGCACCTGGATGAAATCATGGAGATTGAAGAGGAAAAGAAAAGAACTGCCCGAA 840

QY 593 CAGACTACTGGCTACAGCCTGAAATATTGTGAAAAATTATAACCAAGAAACTGGGAGAGA 652

Db 841 CAGACTACTGGCTACAGCCTGAAATATTGTGAAAAATTATAACCAAGAAACTGGGAGAGA 900

QY 653 AATATCATAGAAAAGGCTATTGTTAAGGAGTAATTGACAAATATACAGCTGTTGTGA 712

Db 901 AATATCATAGAAAAGGCTATTGTTAAGGAGTAATTGACAAATATACAGCTGTTGTGA 960

QY 713 AGATGATTGATTCGTGAGACAAGCTGAAACTTCACCAAGCTCATTTAGAGACAGTAATTC 772

Db 961 AGATGATTGATTCGTGAGACAAGCTGAAACTTCACCAAGCTCATTTAGAGACAGTAATTC 1020

QY 773 CAGCACCAGAAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAGAAATGAAGGTACCC 832

Db 1021 CAGCACCAGAAAAGAAATTTCTAGTTTTTAAATGGAGGCTACAGAGAAATGAAGGTACCC 1080

QY 833 TAGATCCATCATGAGAGACTTTTTTCAGCTACTATCGTCATTGAACTGGCCCTTTAA 892

Db 1081 TAGATCCATCATGAGAGACTTTTTTCAGCTACTATCGTCATTGAACTGGCCCTTTAA 1140

QY 893 AAGGACGACAGATTGAAGGAATTCATATGAAGACATTTCTAACTTGCTGAGTTTGA 952

Db 1141 AAGGACGACAGATTGAAGGAATTCATATGAAGACATTTCTAACTTGCTGAGTTTGA 1200

QY 953 AATTCTTTAACAATACCTTTTAAATCTTAAAGCATCAAAATGGTGTCCG 1002

Db 1201 AATTCTTTAACAATA-CATTAAATCTTAAAGCATCAAAATGGTGTCCG 1249

RESULT 6

AX003311

LOCUS AX003311 1002 bp DNA linear PAT 24-AUG-2000

DEFINITION Sequence 4 from Patent WO9929845.

ACCESSION AX003311

VERSION AX003311.1 GI:9927128

KEYWORDS

SOURCE

ORGANISM Homo sapiens (human)

REFERENCE 1

AUTHORS Angulo-Mora, J.F. and Mauffrey, P.

TITLE Sequences coding for kin17 protein and their applications

JOURNAL Patent: WO 9929845-A 4 17-JUN-1999;

ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)

FEATURES

source Location/Qualifiers

1. .1002

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

ORIGIN

Query Match 46.9%; Score 470; DB 6; Length 1002;

Best Local Similarity 100.0%; Pred. No. 8.1e-101;

Matches 470; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 AGTGCACTGAAGCGATAGGAGTTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAGC 504

Db 533 AGTGCACTGAAGCGATAGGAGTTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAGC 592

QY 505 TCAACTCAGTCTTAAGAAAAGAAAAGAAATCTGCATCGATGAAATCATGGAGATT 564

Db 593 TCAACTCAGTCTTAAGAAAAGAAAAGAAATCTGCATCGATGAAATCATGGAGATT 652

QY 565 GAAGAGAAAAGAAAAGAACTGCCCGAACAGACTACTGGCTACAGCCTGAAATTTATG 624

Db 653 GAAGAGAAAAGAAAAGAACTGCCCGAACAGACTACTGGCTACAGCCTGAAATTTATG 712

QY 625 AAAATTTAAACCAAGAACTGGGAGAGAAATATCATAAGAAAAGAAAGCTATTGTTAAGGA 684

Db 713 AAAATTATAACCAAGAACTGGGAGAGAAATATCATAAGAAAAGAAAGCTATTGTTAAGGA 772

QY 685 GTAATTGACAAATATACAGCTGTTGTTGAAGATGATTGATTTCTGGAGACAAGCTGAAACTT 744

Db 773 GTAATTGACAAATATACAGCTGTTGTTGAAGATGATTGATTTCTGGAGACAAGCTGAAACTT 832

QY 745 GACCAGACTCATTTAGAGACAGTAATTCAGCACCAGGAAAAGAAATTTCTAGTTTAAAT 804

Db 833 GACCAGACTCATTTAGAGACAGTAATTCAGCACCAGGAAAAGAAATTTCTAGTTTAAAT 892

QY 805 GGAGGCTACAGAGGAAATGAAGGTACCTTAGAATCCATCAATGAGAAGACTTTTTCAGCT 864

Db 893 GGAGGCTACAGAGGAAATGAAGGTACCTTAGAATCCATCAATGAGAAGACTTTTTCAGCT 952

QY 865 ACTATCGTCATTGAACTGGCCCTTTTAAAGGACGACAGATTGAAAGGAAT 914

Db 953 ACTATCGTCATTGAACTGGCCCTTTTAAAGGACGACAGATTGAAAGGAAT 1002

RESULT 7

CQ722054

LOCUS CQ722054 1117 bp DNA linear PAT 03-FEB-2004

DEFINITION Sequence 7988 from Patent WO02068579.

ACCESSION CQ722054

VERSION CQ722054.1 GI:42282911

KEYWORDS

SOURCE

ORGANISM Homo sapiens (human)

REFERENCE 1

AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.

TITLE Kits, such as nucleic acid arrays, comprising a majority of humanexons or transcripts, for detecting expression and other uses thereof

JOURNAL Patent: WO 02068579-A 7988 06-SEP-2002;

PE Corporation Location/Qualifiers

1. .1117

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

ORIGIN

Query Match 43.7%; Score 438; DB 6; Length 1117;

Best Local Similarity 100.0%; Pred. No. 3.2e-93;

Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 445 AGTGCACTGAAGCGATAGGAGTTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAGC 504

Db 680 AGTGCACTGAAGCGATAGGAGTTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAGC 739

QY 505 TCAACTCAGTCTTAAGAAAAGAAAAGAAATCTGCATCGATGAAATCATGGAGATT 564

Db 740 TCAACTCAGTCTTAAGAAAAGAAAAGAAATCTGCATCGATGAAATCATGGAGATT 799

QY 565 GAAGAGAAAAGAAAAGAACTGCCCGAACAGACTACTGGCTACAGCCTGAAATTTATG 624

Db 800 GAAGAGAAAAGAAAAGAACTGCCCGAACAGACTACTGGCTACAGCCTGAAATTTATG 859

QY 625 AAAATTTAAACCAAGAACTGGGAGAGAAATATCATAAGAAAAGAAAGCTATTGTTAAGGA 684

Db 860 AAAATTTAAACCAAGAACTGGGAGAGAAATATCATAAGAAAAGAAAGCTATTGTTAAGGA 919

QY 685 GTAATTGACAAATATACAGCTGTTGTTGAAGATGATTGATTTCTGGAGACAAGCTGAAACTT 744

Db 920 GTAATTGACAAATATACAGCTGTTGTTGAAGATGATTGATTTCTGGAGACAAGCTGAAACTT 979

QY 745 GACCAGACTCATTTAGAGACAGTAATTCAGCACCAGGAAAAGAAATTTCTAGTTTAAAT 804

Db 980 GACCAGACTCATTTAGAGACAGTAATTCAGCACCAGGAAAAGAAATTTCTAGTTTAAAT 1039

[illegible]

Qy	Db	Match	Similarity	Score	DB %	Length
41	TCGCTCCGTGGTCGCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAACA	829	Conservative	66.0%	0	1.6e-89
3	TACGCTTACGGTTCGCATGGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAACA	829	Conservative	66.0%	0	1.6e-89

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 221547)

Worley,K.C.
Direct Submission
Submitted (01-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 221547)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 19, 2002 this sequence version replaced gi:23611286.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GWGC
Center clone name: CH230-274116
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 208013 bases at least Q40
Consensus quality: 211685 bases at least Q30
Consensus quality: 213907 bases at least Q20
Estimated insert size: 199182; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 198201: contig of 198201 bp in length
* 198202 198301: gap of unknown length
* 198302 199749: contig of 1448 bp in length
* 199750 199849: gap of unknown length
* 199850 201115: contig of 1266 bp in length
* 201116 201215: gap of unknown length
* 201216 202338: contig of 1123 bp in length
* 202339 202439: gap of unknown length
* 202439 203500: contig of 1062 bp in length
* 203501 203600: gap of unknown length
* 203601 204905: contig of 1305 bp in length
* 204906 205005: gap of unknown length
* 205006 206193: contig of 1187 bp in length
* 206193 206292: gap of unknown length
* 206293 207385: contig of 1093 bp in length
* 207386 207486: gap of unknown length
* 207486 208519: contig of 1034 bp in length
* 208520 209759: gap of unknown length
* 209759 209858: gap of unknown length
* 209859 211441: contig of 1583 bp in length
* 211442 211541: gap of unknown length

* 211542 212876: contig of 1335 bp in length
* 212877 212976: gap of unknown length
* 212977 214356: contig of 1280 bp in length
* 214357 214559: gap of unknown length
* 214560 217003: contig of 2647 bp in length
* 217004 217103: gap of unknown length
* 217104 218459: contig of 1356 bp in length
* 218460 218559: gap of unknown length
* 218560 219788: contig of 1229 bp in length
* 219789 219888: gap of unknown length
* 219889 221547: contig of 1659 bp in length.

FEATURES

source
1..221547
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-274116"
1..1272
/note="wgs end extension
clone_end:Sp6"
misc_feature
/complement(4416..5263)
/note="clone boundary
clone_end:Sp6
site:
end_sequence:BZ266352"
192445..193553
/note="wgs_contig"
195743..196592
/note="clone boundary
clone_end:t7
site:
end_sequence:BZ266349"
198202..198301
/estimated_length=unknown
199750..199849
/estimated_length=unknown
201116..201215
/estimated_length=unknown
202339..202438
/estimated_length=unknown
203501..203600
/estimated_length=unknown
204906..205005
/estimated_length=unknown
206193..206292
/estimated_length=unknown
207386..207485
/estimated_length=unknown
208520..208619
/estimated_length=unknown
209759..209858
/estimated_length=unknown
211442..211541
/estimated_length=unknown
212877..212976
/estimated_length=unknown
214257..214356
/estimated_length=unknown
217004..217103
/estimated_length=unknown
218460..218559
/estimated_length=unknown
219789..219888
/estimated_length=unknown
/estimated_length=unknown

Query Match 42.2%; Score 422.4; DB 14; Length 221547;

Best Local Similarity 66.0%; Pred No. 1.2e-89;
Matches 829; Conservative 0; Mismatches 126; Indels 301; Gaps 3;

QY 41 TCGCTGCCGTGCGCCATGGGAAAGTCGATTTCTTACTCCCAAGGCTATCGCCAACA 100

DB 212995 TAGCGCTCAGGGTCGCGCAAGTGGCAAGTCGATTTCTTCTGAGCCCAAGCCATCGCCAATA 213054


```
QY 101 GGATCAAGTCCAGGGGCTCGCAAGCTAGCTGGTATTGCCAGATGTGCCAGAACGCT 160
Db 213055 GAATTAAAGTCCAAAGGGCTCCAGAGCTTCGCTGGTACTGCCAGATGTGCCAAAGCAAT 213114
QY 161 GCCGGACGAGAAATGGCTTTAAAGTGCATTTGTATGTCCGAATCTCATCAGAGACAACACTAT 220
Db 213115 GCCCGCAGCAGAAATGGCTTTAAAGTGCATCTGTATGTCTGAATCTCATCAAGAGACAACACTGT 213174
QY 221 TGTGGCTTCAGAAATCTCTCAGCAGTTCATGGATTTATTTTCAGAGGAATTCGGAATG 280
Db 213175 TGTGGCTTCAGAAATCCCTCAGCAGTTCATGGATTTATTTTCAGAGGAATTCGGAATG 213234
QY 281 ACTTTCTAGAATCTCTCAGAGAGCTTTGGCAGCTAAAGGGTCCCAACAACATTTGTCT 340
Db 213235 ACTTTCTGGAACTCTGAGCGAGCTTTGGCAGCTAAAGGGTCCCAACAACATTTGTCT 213294
QY 341 ACAACGAATACATCAGCCACCGAGAGCACAATCCACATGAATGCCATCTCAGTGGGAACCTC 400
Db 213295 ACAATGAATACATCAGCCACCGAGAGCACAATCCACATGAACGCTACTCAGTGGGAGACAC 213354
QY 401 TGACTGATTTTACTAAGTGTGGCAGAGAGAGCTTTGTGCAA----- 443
Db 213355 TGACCGACTTTTACCAAGTGGCTGGCAGAGAGGGCTTTGTGTAAGTGGATGAGACACCGA 213414
QY 444 ----- 443
Db 213415 AAGGCTGGTACATTCAGTACATAGACAGAGACCCAGAAACCCTCGTCGGCAACTGGAT 213474
QY 444 ----- 443
Db 213475 TAGAAAAAAGNAGNAGNAGNAGTCTGGACGATGAAGAAAAAACTGCCAAGTTCATTGAGG 213534
QY 444 ----- 443
Db 213535 AGCAGGTGAGAAGAGGGCTTGAAGGGAAAGAGCAGCAGAGACACCTGTTTTTACAGAACTAA 213594
QY 444 ----- 443
Db 213595 GCCGAGAAATGAGGAAGAAAAAGTTACGTTCAATCTGAATTAAGAGCGGGTGGCTCAG 213654
QY 444 -----AAGTGCACTGAAGACGATAGGAA 466
Db 213655 CGGGAGCTACAACATCCAAAGTCAAGCTCTTTGGGACCAAGTGCACTGAAGCTGTGGGA 213714
QY 467 GTTCAGCATCAGTGAAGCGAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGA 526
Db 213715 GCGCAGCATCCGGGAAACGGAAGAGTCTTCAAGAGCTCCGCCACGCT-----GCGA 213769
QY 527 AGAAAAAAGAAATCTGCACTGGTGAATCATGGAGATTGAAGGAAAGAAAGAACTG 586
Db 213769 AGAAGAAAGATCCGCCCTGGATGAGATCATGGAGCTCGAAGAGGAAAGAAAGAAAGACCG 213829
QY 587 CCCGAACAGACTACTCGCTACAGCTGAAATTTATTTGAAAAATTAACCAAGAAACTGG 646
Db 213829 CACGACAGACGCTGGTTACAGCGGGGATCGTTGTGAAAAATTAACGAGAACTTG 213888
QY 647 GAGAGAAATATCATAGAAAAAGGCTATTTGTAAGGAAGTAATGACAAATATACAGCTG 706
Db 213889 GGGAGAAATATCATAGAAAAAGGCTCGTTAAGGAAGTGAATGACAGGTTACACAGCTG 213948
QY 707 TTGTGAAGATGATTTCTGGAGACAGCTGAACTTGACCACTCATTTAGAGACAG 766
Db 213949 TGGTAAAGATGACTGACTCTGGAGACAGGCTGAACTGGACCACTCATTTAGAGACAG 214008
QY 767 TAAATCCAGCACAGGAAAAAGAAATTCATGTTTTTAAATGAGGCTACAGAGGAATGAAG 826
Db 214009 TCATTCGGCCCCGGGAAAAAGGGTTCTAGTTTTTAATGAGGCTACAGAGGAATGAAG 214068
QY 827 GTACCCTAGAATCCATCAATGAGAAGACTTTTTCAGCTACTATCGTCAATTGAAACTGGCC 886
Db 214069 GCACCTCGAATCCATCAATGAGAAGGCTTTTTCAGCCACGATAGTCAATTGAAACTGGAC 214128
QY 887 CTTTAAAGGACGCGAGAGTTGAGGAATTCATATGAAGACATTTCTAAACTTGCCTGAG 946
```

```
Db 214129 CTTTGAAGAGCGCAGAGTTGAAGGTATTCAATATGAAGACATATCTAAACTTGTCTGAG 214188
QY 947 TTTGAAATTTTGAACATATACCTTTAAATCTTTAAAGCATCAAAATTTGTTTCG 1002
Db 214189 TTTGAAATTTTGAATACCAACATTCGAAACT-GTGAAGCATCAAAATTTGATTAGC 214243

RESULT 11
AX003331 1390 bp DNA linear PAT 24-AUG-2000
LOCUS Sequence 24 from Patent WO9929845.
DEFINITION AX003331
ACCESSION AX003331
VERSION AX003331.1 GI:9927146
KEYWORDS Mus sp.
SOURCE Mus sp.
ORGANISM Mus sp.
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
  Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Angulo-Mora, J.P. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 24 17-JUN-1999;
  ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
  (FR)
FEATURES             Location/Qualifiers
     source            1..1390
                        /organism="Mus sp."
                        /mol_type="unassigned DNA"
                        /db_xref="taxon:10095"
ORIGIN
Query Match          41.6%; Score 416.6; DB 6; Length 1390;
Best Local Similarity 66.1%; Pred. No. 3.9e-88;
Matches 819; Conservative 0; Mismatches 119; Indels 301; Gaps 3;

QY 58 ATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATGCCAAGCAGTGCAGAGAGTGCAGAGTGC 117
Db 1 ATGGGCAAGTCGGATTTTCTGAGCCCAAGGCCATGCCCAATAGATTAAAGTCCAAAGGG 60
QY 118 CTCAGAGAGCTACGCTGGTATTCAGAGATGTGCCAGAGCAGTGCAGAGAGTGCAGAGTGC 177
Db 61 CTCAGAGAGCTTCGCTGGTACTGCCAGATGTGCCAAGCAATGCCGCGCAGAGATGCG 120
QY 178 TTTAAGTGCATTTGTATGTCGGAATCTCATCAGAGACAACTATTTGCTGGCTTCAGAAAT 237
Db 121 TTTAAGTGCATCTGTATGTCTGAATCTCATCAAGAGCAACTGTGCTGGCTTCAGAAAC 180
QY 238 CCTCAGCAGTTTATGGATTTATTTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTC 297
Db 181 CCTCAGCAGTTTATGGATTTATTTTTCAGAGGAATTCGAAATGACTTTCTGGAACCTCTG 240
QY 298 AGGAGAGCGCTTTGGCACTAAAGGGTCCACAAACAACTGCTACACGAAATCATCAGC 357
Db 241 AGCGAGCGCTTTGGCACTAAAGGGTCCACAAACAACTGCTACAAATGAATATCATCAGC 300
QY 358 CACCGAGAGCAGATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTAAG 417
Db 301 CACCGAGAGCAGATCCACATGAACGCTACCAGTGGGAGACACTGACCGGACTTTTACCAG 360
QY 418 TGCTGGGAGAGAGGCTTTGTGCAA----- 443
Db 361 TGGCTGGGAGAGAGGCTTTGTGTAAGTGGATGAGACACCGAAAGGCTGATACATTGAG 420
QY 444 ----- 443
Db 421 TACATAGACAGAGACCCAGAAACCATCCGTCGGCAACTGGAATTTAGAAAAAGAGAAAG 480
QY 444 ----- 443
Db 481 CAAGATCTGGACGATGAAGAAAAAACTGCCAAGTTTCATTGAGGAGCAGGTGAGAAAGGC 540
```

Qy	444	-----	442
Db	541	CTGGAAGGAAAGACGACGAGACACCTGTTTTCACAGAACTTAGCCGAGAAAATCAGGAA	600
Qy	444	-----	443
Db	601	GA AAAAGTTTACGTTCAATCTGAAATAAAGGACGGGTGGCTCAGCGGAGCTTACAACATCC	660
Qy	444	-----	442
Db	661	AAGTCAAGCTCTTTGGGACCCAAAGTGCACTGNAAGCTGCTGGGGAGCGCAGCATCCGGGAAA	720
Qy	484	CGAAAAAGATCTTCCAGAGCTCAACTCAGTCTTAAGAAAAAGAGAAAAAGAAATCTGCA	543
Db	721	CGGAAAGAGTCTTACACAGAGCTCCGCCAGCT-----GCGAAGAAAGAAAGTCCGCC	774
Qy	544	CTGGATGAATCATGGAGATTGAAGAGGAAAAAGAAAGAACTGCCGACAGACTACTGG	603
Db	775	CTGGATGAGATCATGGAGCTCGAAGAGGAAAAAGAAAGGACCGCACGACAGACGCTCG	833
Qy	604	CTACAGCTCGAAATTAATTGTGAAAAATTATTAACCAAGAAACTGGGAGAGAAAAATATCATAAG	663
Db	835	TTACAGCGCGGGATCGTTGTGAAAAATTATTAACGAAAGAACTTGGGAGAGAAATATCACAG	894
Qy	664	AAAAAGGCTATTGTTAAGGAAGTAATTAACAAAATATACAGCTGTTGTAAGATGATGAT	723
Db	895	AAGAAAGGGCTCGTTAAGGAAGTGATTTGACAGGTACACAGCTGTGGTAAAGATGACTGAC	954
Qy	724	TCTGGAGACAAGCTGAAACTTTGACCAAGACTCATTTAGAGACAGTAATTCAGACACGAGA	783
Db	955	TCTGGAGACAAGCTGAAACTGGACCAGACTCATTTAGAGACAGTCAATTCGGGCCCGGGG	1014
Qy	784	AAAAAGAAATCTAGTTTTAAATGGAGGCTACACAGAGAAATGAAGTACCCCTAGAAATCCATC	843
Db	1015	AAAAGGGTCTAGTTTTAAATGGAGGCTACACAGAGAAATGAAGCACTCTCGAATCCATC	1074
Qy	844	AATGAGAGAGCTTTTTCAGCTACTATCGTCAATGAAACTGCGCCCTTTAAAGAGCGCAGA	903
Db	1075	AATGAGAGAGCTTTTTCAGCCACGATAGTCAATGAAACTGAGACCTTTGAAAGGAGCGAGA	1134
Qy	904	GTTGAAAGAAATCAATATGAAGACATTTCTAACTTCGCTGAGTTTGAAATTTGTTAAC	963
Db	1135	GTTGAAAGGATTTCAATATGAAGACATATCTAAACTTCGTTGAGTTTGAAATTTGATAAC	1194
Qy	964	AATACCTTTTAAATCTTTAAAGCATCAAAATGGTGTTCGC	1002
Db	1195	AACACATGAACT-GTGAGCATCAAAATGGTGTGTTAGC	1232
RESULT 12			
BC058169			
LOCUS			
DEFINITION	BC058169	2710 bp mRNA linear ROD 13-FEB-2004	
ACCESSION		Mus musculus antigenic determinant of rec-A protein, mRNA (cdna	
VERSION		clone IMAGE:5026202), partial cds.	
KEYWORDS		BC058169	
SOURCE		BC058169.1 GI:37194909	
ORGANISM		Mus musculus (house mouse)	
REFERENCE			
AUTHORS			

```

Db      2189  ATGGAGCTCGAAGGAGAAAGAAAGGACCGCAGCAGCAGCAGCCTGGTTACAGCCGGGG 2248
Qy      616  ATTATTGTGAAAATTATAACCAAGAAACTCGGAGAGAGAAATATCATTAAGAAAAGGCTATT 675
Db      2249  ATCGTTGTGAAAATTATAACCAAGAAAGCTTTGGGGAGAGAAATATCACAGAGAAAGGGGTC 2308
Qy      676  GTTAAAGGAAGCTAATTCAGCAATATACAGCTGTGTGTGAAGATGATTGATCTCGAGACAAAG 735
Db      2309  GTTAAAGGAAGTGATTGACAGGTACACAGCTGTGTGTGAAGATGACTGACTCTCGAGACAGG 2368
Qy      736  CTGAAACTTCACAGAGCTCAATTTAGAGACAGTAATTTCCAGCACCAGCAAGAAAAGAAATCTTA 795
Db      2369  CTGAACTGACAGAGCTCAATTTAGAGACAGTCAATTCGCGCCCGGGGAAAGGGTTCTTA 2428
Qy      796  GTTTTAAATCGAGGCTACAGAGGAATGAAGGTACCTAGAAATCCATCATATGAGAGACT 855
Db      2429  GTTTTAAATCGAGGCTACAGAGGAATGAAGGCACTCTCGAATCCATCATATGAGAGGCT 2488
Qy      856  TTTTCAGCTACTATCGTCAATTTGAAACTGGCCCTTTAAAGGAGCGCAGAGTTGAAGGAATT 915
Db      2489  TTTTCAGCCAGTAGTCATTTGAAACTGGACCTTTGAAGAGACGACAGTTGAAGGTATT 2548
Qy      916  CAATATGAGACATTTCTAAACTTGCCTGAGTTTGAAGAAATTTGTTAAACATACCTTTAAA 975
Db      2549  CAATATGAGACATATCTAACTTGTGATTTGAAATTTGATTAACACACATTTGAAA 2608
Qy      976  ATCTTAAAGCATCAAAATGCGTTCGC 1002
Db      2609  CT-CTGAAGCATCAAAATGCGTTCG 2634

RESULT 13
AC114615/c
LOCUS   AC114615 180882 bp DNA linear HTG 09-MAR-2004
DEFINITION Mus musculus clone RP24-82119, WORKING DRAFT SEQUENCE, 4 ordered
          pieces.
ACCESSION AC114615
VERSION   AC114615.5 GI:44886278
KEYWORDS  HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
          Sciurognathi; Muroidae; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 180882)
AUTHORS  Birren,B., Nusbaum,C. and Lander,E.
TITLE    Mus musculus, clone RP24-82119
JOURNAL  Unpublished
FEATURES  Feature Table
          2 (bases 1 to 180882)
          Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
          Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
          Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
          Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
          Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
          Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
          Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
          Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Kanat,A., Karatas,A., Kells,C., Larocque,K., Lamazares,R.,
          Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
          MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
          McCarthy,W., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
          Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
          Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
          Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
          Raymond,C., Retta,R., Riback,M., Riley,R., Rise,C., Rogov,P.,
          Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
          Seaman,J., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
          Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
          Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
          Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
          Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
          Direct Submission
          Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome

```

```

Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 180882)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collamore,A., Cook,A., Cooke,P., Dorris,L., Erickson,J., Faro,S.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrum,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-MAR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2004 this sequence version replaced gi:24182248.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24315
Center clone name: 82_1_19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191018 bases at least Q40
Consensus quality: 191303 bases at least Q30
Consensus quality: 191388 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 191574; sum-of-contigs
Quality coverage: 12.2 in Q20 bases; agarose-fp
Quality coverage: 11.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 89963: contig of 89963 bp in length
* 89964 90063: gap of 100 bp
* 90064 94560: contig of 4497 bp in length
* 94561 94660: gap of 100 bp
* 94661 169601: contig of 74941 bp in length
* 169602 169701: gap of 100 bp
* 169702 180882: contig of 1181 bp in length.
FEATURES             Location/Qualifiers
     source            1..180882
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP24-82119"
                     /clone_lib="RPC1-24 Male Mouse BAC"
misc_feature          1..89963

```

/note="assembly_fragment clone_end:SP6 vector_side:left" 89964..90063 /estimated_length=100 90064..94560 /note="assembly_fragment" 94561..94660 /estimated_length=100 94661..169601 /note="assembly_fragment" 169602..169701 /estimated_length=100 169702..180882 /note="assembly_fragment clone_end:T7 vector_side:right"		39.0%; Score 390.6; DB 14; Length 180882; Best Local Similarity 83.1%; Pred. No. 4.5e-82; Matches 471; Conservative 0; Mismatches 89; Indels 7; Gaps 2;	
gap	436 TTGTGCAAAAGTGCACCTGAAGCAGTAGGAAGTTTCAGCATTCAGTGAACGAAAGAAATCT	495	
misc_feature	46568 TTGGGACCAAGTGCATCTGAAGCTCTGGGGAGCGCAGCATCCGGGAAACGGAAAGAGTCT	46509	
gap	496 TCCGAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCACCTGGATGAAATC	555	
misc_feature	46508 TCACAGAGCTCCGCCAGCTGC-----GAAGAAGAAGAGTCCGCCCTGGATGAGATC	46455	
gap	556 ATCGAGATTGAAGAGAAAAAGAAAGAACTGCCGCAACAGACTACTTGCTACAGCCTGAA	615	
misc_feature	46454 ATCGAGCTCGAAGAGAAAAAGAAAGGACCGCACGACAGACGCTGGTTACAGCCGGGG	46395	
gap	616 ATTATTGTGAAATTATACCAAGAACTGGGAGAGAAATATCATAGAAGAAAGCGTATT	675	
misc_feature	46394 ATCGTTGTGAAATTTATAACGAAGAAGCTTGGGGAGAAATATCACAAGAAGAAAGGGTC	46335	
gap	676 GTTAAGGAAGTAAATTCACAAATATACAGCTGTTGTGAAGATGATTGATTCTGGAGACAAG	735	
misc_feature	46334 GTTAAGGAAGTGATTGACAGGTACACAGCTGTGTTAAGATGACTGACTCTGGAGACAGG	46275	
gap	736 CTGAAACTTGACAGACTCATTTAGAGACAGTAATTTCCAGCACGACGAGAAAAAGAAATCTTA	795	
misc_feature	46274 CTGAAACTGGACAGACTCATTTAGAGACAGTCAATTCGGGCCCGCGGAAAAAGGGTTCTA	46215	
gap	796 GTTTTAAATGGAGGCTACAGAGAAATGAAGTACCCTAGATCCATCAATCAGAGACT	855	
misc_feature	46214 GTTTTAAATGGAGGCTACAGAGAAATGAAGGCATCTCGAATCCATCAATGAGAAGGCT	46155	
gap	856 TTTTTCAGCTACTATCGTCAATTGAAACTGGCCCTTTAAAAAGGACGACAGAGTTTGAAGGAATT	915	
misc_feature	46154 TTTTTCAGCCACGATAGTCAATTGAACTGACCTTTGAAGAGCGCAGAGTTTGAAGGTAAT	46095	
gap	916 CAATATGAAGACATTTCTAAACTTGCTCAGTTTGAAAATTTGTTTAAACAATACCTTTAAA	975	
misc_feature	46094 CAATATGAAGACATATCTAAACTTGCTTGAGTTTGAAAATTTGATTAACAACACATTGAAA	46035	
gap	976 ATCTTAAAGCATCAAAATGGTGTTCGC	1002	
misc_feature	46034 CT-GTGAAGCATCAAAATGGTGTTAGC	46009	
RESULT 14			
AC124010			
LOCUS	Mus musculus chromosome UNK clone RP23-119N4, WORKING DRAFT		
DEFINITION	SEQUENCE, 6 unordered pieces.		
ACCESSION	AC124010		
VERSION	AC124010.2 GI:22475963		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Mus musculus (house mouse)		

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 182400)
TITLE	McPherson,J.D. and Waterston,R.H.
JOURNAL	The sequence of Mus musculus clone Unpublished
REFERENCE	2 (bases 1 to 182400)
AUTHORS	McPherson,J.D. and Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (06-JUN-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE	3 (bases 1 to 182400)
AUTHORS	McPherson,J.D. and Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	On Aug 25, 2002 this sequence version replaced gi:21327639.
----- Genome Center -----	
Center: Washington University Genome Sequencing Center	
Center code: WUGSC	
Web site: http://genome.wustl.edu/gsc/index.shtml	
Contact: submissions@watson.wustl.edu	
----- Project Information -----	
Center project name: M_BA0119N04	
----- Summary Statistics -----	
Sequencing vector: M13; 0%	
Chemistry: Dye-primer ET; 0% of reads	
Assembly program: Phrap; version 0.990319	
Consensus quality: 183528 bases at least Q40	
Consensus quality: 184210 bases at least Q30	
Consensus quality: 184756 bases at least Q20	
Insert size: 197000; agarose-fp	
Quality coverage: 9.53 in Q20 bases; agarose-fp	
Quality coverage: 9.61 in Q20 bases; sum-of-contigs	
* NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
* 1 11868: contig of 11868 bp in length	
* 11869 11968: gap of unknown length	
* 11969 28906: contig of 16938 bp in length	
* 28907 29006: gap of unknown length	
* 29007 72773: contig of 43767 bp in length	
* 72774 72874: gap of unknown length	
* 72874 124926: contig of 52053 bp in length	
* 124927 125026: gap of unknown length	
* 125027 181902: contig of 56876 bp in length	
* 181903 182002: gap of unknown length	
* 182003 182400: contig of 398 bp in length.	
Location/Qualifiers	
1..182400	
/organism="Mus musculus"	
/mol_type="genomic DNA"	
/db_xref="taxon:10090"	
/chromosome="UNK"	
/clone="RP23-119N4"	
1..11868	
/note="assembly_name:Contig86"	
11869..11968	
/estimated_length=unknown	
11869..28906	
misc_feature	
gap	
misc_feature	

```

/note="assembly_name:Contig87"
28907..29006
/estimated_length=unknown
misc_feature
29007..72773
/note="assembly_name:Contig88"
72774..72873
/estimated_length=unknown
gap
72874..124926
/note="assembly_name:Contig89"
124927..125026
/estimated_length=unknown
misc_feature
125027..181902
/note="assembly_name:Contig90"
181903..182002
/estimated_length=unknown
gap
182003..182400
/note="assembly_name:Contig16"

```

ORIGIN

```

Query Match      39.0%; Score 390.6; DB 14; Length 182400;
Best Local Similarity 83.1%; Pred. No. 4.5e-82;
Matches 471; Conservative 0; Mismatches 89; Indels 7; Gaps 2;

Qy 436 TTGTGCAAAAGTCACCTGAAGACGATAGGAAGTTTCAGCATCAGTGAACCGAAAAGATCT 495
Db 168845 TTGGGACCAAGTCACCTGAAGCTGCTGGGGAGCGCAGCATCCGGGAAACCGAAAGAGTCT 168904

Qy 496 TCCAGAGCTCACTCACTAGTCTTAAGAAAAGAAAAGAAATCTGCACCTGGATGAATC 555
Db 168905 TCACAGAGCTCGGCCAGCCTGC-----GAAGAAGAAAGAGTCCGGCTGGATGAGATC 168958

Qy 556 ATGAGATTGAGAGGAAAGAAAAGAACTGCCCGACAGACACTCTGCTACAGCCTGAA 615
Db 168959 ATGAGACTCGAGAGGAAAAGAAAAGGACCGCACAGACGCTGCTGTTACAGCCGGG 169018

Qy 616 ATTATTGTGAAAATTATACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCTATT 675
Db 169019 ATCGTTGTGAAAATTATACGAGAGAGCTTGGGAGAAATATCATAGAAAAGGCTC 169078

Qy 676 GTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATGATTCGGAGACAAG 735
Db 169079 GTTAAGGAAGTAATTGACAGGTACACAGCTGTGTGTAAGATGACTGACTCTGGAGACAG 169138

Qy 736 CTGAAACTTGCACAGACTATTAGACAGCAATATTCAGCACAGGAAAAGAAATCTTA 795
Db 169139 CTGAAACTTGCACAGACTATTAGACAGCAATATTCAGCACAGCTATTCGGCCCGGAAAAGGTTCTA 169198

Qy 796 GTTTTAAATGAGGCTACAGAGGAAATGAAGTACCCTAGAAATCCATCAATGAGAGACT 855
Db 169199 GTTTTAAATGAGGCTACAGAGGAAATGAAGGCACTCTCGAATCCATCAATGAGAGGCT 169258

Qy 856 TTTTCAGCTACTATCGTCAATGAACCTGGCCCTTTAAAGGACGACAGATTGGAAGGAAT 915
Db 169259 TTTTCAGCCAGTAGTCAATGAACTGGACCTTTGAAAGGACGACAGATTGGAAGGATT 169318

Qy 916 CAATATGAAGACATTTCTTAACCTGCTGAGTTTGAAATTTGTTAACAATACCTTTAA 975
Db 169319 CAATATGAAGACATATCTAAACTTGTCTGAGTTTGAAATTTGTTAACAACATTTGAAA 169378

Qy 976 ATCTTAAGCATCAAAATTTGGTTTCGC 1002
Db 169379 CT-GTGAAGCATCAAAATTTGGTTTAGC 169404

```

```

RESULT 15
LOCUS AL772367/c 185378 bp DNA linear ROD 02-SEP-2002
DEFINITION Mouse DNA sequence from clone RP23-119N4 on chromosome 2, complete
sequence.
ACCESSION AL772367
VERSION AL772367.7 GI:22759512
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)

```

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
1 (bases 1 to 185378)
Wood, J.

REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
Submitted (30-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Sep 9, 2002 this sequence version replaced gi:2265488.

COMMENT

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-119N4 is
constructed by the group of Pieter de Jong.

For further details see http://www.chori.org/bacpac/home.htm

FEATURES
source

```

Location/Qualifiers
1..185378
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone_lib="RP23-119N4"
/clone_lib="RPCI-23"

```

ORIGIN

```

Query Match      39.0%; Score 390.6; DB 9; Length 185378;
Best Local Similarity 83.1%; Pred. No. 4.5e-82;
Matches 471; Conservative 0; Mismatches 89; Indels 7; Gaps 2;

Qy 436 TTGTGCAAAAGTCACCTGAAGACGATAGGAAGTTTCAGCATCAGTGAACCGAAAAGATCT 495
Db 139982 TTGGGACCAAGTCACCTGAAGCTGCTGGGGAGCGCAGCATCCGGGAAACCGAAAGAGTCT 139923

Qy 496 TCCAGAGCTCACTCACTAGTCTTAAGAAAAGAAAAGAAATCTGCACCTGGATGAATC 555
Db 139922 TCACAGAGCTCGGCCAGCCTGC-----GAAGAAGAAAGAGTCCGGCTGGATGAGATC 139869

Qy 556 ATGAGATTGAGAGGAAAGAAAAGAACTGCCCGACAGACACTCTGCTACAGCCTGAA 615
Db 139868 ATGAGCTCGAAGAGGAAAAGAAAAGGACCGCACAGAGCGCTGTTACAGCCGGG 139809

Qy 616 ATTATTGTGAAAATTATACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCTATT 675
Db 139808 ATCGTTGTGAAAATTATACGAGAGAGCTTGGGAGAAATATCATAGAAAAGGCTC 139749

Qy 676 GTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATGATTCGGAGACAAG 735
Db 139748 GTTAAGGAAGTAATTGACAGGTACACAGCTGTGTGTAAGATGACTGACTCTGGAGACAG 139689

```

QY	736	CTGAAACTTGACCAGACTCATTTAGAGACAGTAATTCACGACCACGAGAAAAGAATTCTA	795
Db	139688		
QY	796	GTGAAACTGGACCAGACTCATTTAGAGACAGTCAITTCGGCCCCCGGGGAAAAGGGTTCTA	139629
Db	139628		
QY	856	TTTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCTAGAACCTCCATCAATGAGAAGACT	855
Db	139568		
QY	916	TTTTTAAATGGAGGCTACAGAGGAAATGAAGGCACTCTCGATCCATCAATGAGAAGGCT	139569
Db	139509		
QY	976	ATCTTAAAGCATCAAATTTGGTTTCGC	1002
Db	139448	CT-GTGAAGCATCAAATTTGGTTTAGC	139423

Search completed: November 26, 2005, 19:10:59
Job time : 5335.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 642.701 Seconds
(without alignments)
10390.557 Million cell updates/sec

Title: US-09-555-529-3
Perfect score: 1002
Sequence: 1 tgattcagctggtaccgc.....agcatcaattggtgttcgc 1002

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_21:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*
14: Geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1002	100.0	1002	2	Aax85551 CDNA of a
2	704.6	70.3	1102	2	Aax85550 CDNA of a
3	696.4	69.5	1296	2	Aax85549 CDNA of a
4	470	46.9	1002	2	Aax85552 Probe der
5	426.4	42.6	679	13	Adq56943 Novel car
6	416.6	41.6	1390	2	Aax85570 CDNA of a
7	415.2	41.4	1458	2	Aaq79936 Murine Ki
8	267.2	26.7	591	4	Aah34095 Human col
9	201.4	20.1	1372	12	Adp22451 Sea-squi
10	176	17.6	1713	13	Adc17629 Plant CD
11	176	17.6	1780	3	Aax52591 CDNA enco
12	176	17.6	1780	3	Aax52590 CDNA enco
13	171	17.1	300	3	Aaa01559 Human col
14	170.8	17.0	1614	13	Ado84255 Plant fu
15	170.8	17.0	1713	13	Adx47874 Plant fu
16	170.8	17.0	1764	3	Aax52589 CDNA enco
17	168.2	16.8	1459	3	Aac47248 Arabidops
18	163.4	16.3	1241	4	Abi29199 Drosophil
19	163.4	16.3	3313	4	Abi29198 Drosophil

20	163.4	16.3	4394	4	ABL29184
21	163.4	16.3	4582	4	ABL18194
22	136.4	13.6	1032	3	AAC48794
23	126.4	12.6	549	3	AAC94929
24	100.2	10.0	270	6	ABL73715
25	97.2	9.7	398	13	ACF87901
26	94.2	9.4	696	3	AAC38477
27	77.4	7.7	431	3	AAC48801
28	60	6.0	60	6	ABN33592
29	56.8	5.7	2985	5	ABL29210
30	54.4	5.4	65	6	ABN51894
31	50.6	5.0	537	13	ACN53339
32	50	5.0	778	6	ABQ15588
33	50	5.0	778	6	ABQ15589
34	48.2	4.8	2000	11	ACL35887
35	48	4.8	2000	8	ADA71938
36	47.4	4.7	5940	3	AAx70105
37	46.2	4.6	9964	6	ABL32098
38	45.2	4.5	461	13	ACN55894
39	44.6	4.4	4316	10	ADB54224
40	44.6	4.4	4316	13	ADN89522
41	44.6	4.4	6182	6	ABL34015
42	44	4.4	6821	3	AAx70188
43	43.8	4.4	16888	4	AAx46556
44	43.6	4.4	529	10	ACC42649
45	43.6	4.4	612	14	ACL64894

ALIGNMENTS

RESULT 1
AAx85551
ID AAx85551 standard; cDNA; 1002 BP.
XX
AC AAx85551;
XX
DT 07-SBP-1999 (first entry)
XX
DE cDNA of a gene coding for a truncated human kin17 protein.
XX
KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW protein interaction; curved DNA; HIV replication; HIV integration;
KW repair enzyme; ss.
OS Homo sapiens.
XX
XX FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
XX 09-DEC-1997; 97FR-00015536.
XX
PR (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PA Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX
FI WPI; 1999-359999/31.
XX
DR New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.
XX
PS Claim 5; Page 32; 69pp; French.
XX
CC The present sequence encodes a truncated human kin17 protein with amino
CC acids 129-228 deleted. The mammalian kin17 protein is useful for
CC preparing a medicament for controlling cell proliferation or for
CC controlling fertility. The medicaments can also be used to treat
CC hyperproliferative diseases. Fragments between amino acids 55 and 235
CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein
CC are useful for regulating the interaction between proteins and curved

CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
CC proliferation
XX
SQ Sequence 1002 BP; 343 A; 189 C; 230 G; 240 T; 0 U; 0 Other;

Query Match 100.0%; Score 1002; DB 2; Length 1002;
Best Local Similarity 100.0%; Pred. No. 1.5e-265;
Matches 1002; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGATTCGAGCTCGGTACCGGGGATCCGATTAGAAAAGTGCCTGCGGTGTCGCATG 60
DB 1 TGATTCGAGCTCGGTACCGGGGATCCGATTAGAAAAGTGCCTGCGGTGTCGCATG 60

QY 61 GGGAGTCGGATTTCTTACTCCCAAGGCTATGCCAACAGGATCAAGTCCCAAGGGGCTG 120
DB 61 GGGAGTCGGATTTCTTACTCCCAAGGCTATGCCAACAGGATCAAGTCCCAAGGGGCTG 120

QY 121 CAGAAGCTACGCTGATTTGCCAGATGTGCCAGAACAGTGCCTGGGACGAGAAATGCTTT 180
DB 121 CAGAAGCTACGCTGATTTGCCAGATGTGCCAGAACAGTGCCTGGGACGAGAAATGCTTT 180

QY 181 AAGTGTCAATGTATGTCGGAATCTCATCAGAGCAACTATTGCTGCTTCAGAAAATCCT 240
DB 181 AAGTGTCAATGTATGTCGGAATCTCATCAGAGCAACTATTGCTGCTTCAGAAAATCCT 240

QY 241 CAGCAGTTATGGATTTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTTCTCAGG 300
DB 241 CAGCAGTTATGGATTTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTTCTCAGG 300

QY 301 AGACGCTTTGGCACTAAAAGGTCCTCAACAACATTTGCTCAACGAATACATCAGCCAC 360
DB 301 AGACGCTTTGGCACTAAAAGGTCCTCAACAACATTTGCTCAACGAATACATCAGCCAC 360

QY 361 CGAGAGCATCCACATGAATGCCACTCAGTGGGAACTCTGACTATTTTACTAAGTGG 420
DB 361 CGAGAGCATCCACATGAATGCCACTCAGTGGGAACTCTGACTATTTTACTAAGTGG 420

QY 421 CTGGGCAGAGAAGCTTTGTGCAAAAGTCACCTGAAGACGATAGGAAGTTTCAGCATCAGTG 480
DB 421 CTGGGCAGAGAAGCTTTGTGCAAAAGTCACCTGAAGACGATAGGAAGTTTCAGCATCAGTG 480

QY 481 AAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAAGAAATCT 540
DB 481 AAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAGAAAAGAAATCT 540

QY 541 GCACCTGATGAATCATGTGAGATTGAAGAGGAAAGAAAGAACTGCCCGAACAGACTAC 600
DB 541 GCACCTGATGAATCATGTGAGATTGAAGAGGAAAGAAAGAACTGCCCGAACAGACTAC 600

QY 601 TGCTACAGCTGAAATTTATTTGAAAATTTATACCAAGAACTGGGAGAGAAATATCAT 660
DB 601 TGCTACAGCTGAAATTTATTTGAAAATTTATACCAAGAACTGGGAGAGAAATATCAT 660

QY 661 AAGAAAAGCTTATTTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAAGATGTT 720
DB 661 AAGAAAAGCTTATTTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAAGATGTT 720

QY 721 GATTCCTGGAGACAAGCTGAAACTTGCACAGACTCATTTAGAGACAGTAATTCAGCACCA 780
DB 721 GATTCCTGGAGACAAGCTGAAACTTGCACAGACTCATTTAGAGACAGTAATTCAGCACCA 780

QY 781 GGAAGAAAGAAATCTAGTTTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAATCC 840
DB 781 GGAAGAAAGAAATCTAGTTTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAATCC 840

QY 841 ATCAATGAGAAGACTTTTTCAGCTACTATCGTCAATGAAACTGGCCCTTTAAAGAACGC 900
DB 841 ATCAATGAGAAGACTTTTTCAGCTACTATCGTCAATGAAACTGGCCCTTTAAAGAACGC 900

QY 901 AGAGTTTGAAGGAATTCAAATATGAAGACATTTCTTAAACTTCGCTGAGTTTGAAAATTTGTT 960
DB 901 AGAGTTTGAAGGAATTCAAATATGAAGACATTTCTTAAACTTCGCTGAGTTTGAAAATTTGTT 960

DB 901 AGAGTTTGAAGGAATTCAAATATGAAGACATTTCTTAAACTTCGCTGAGTTTGAAAATTTGTT 960
QY 961 AACATACCTTTAAATCTTTAAAGCATCAAAATTTGTTGTCGC 1002
DB 961 AACATACCTTTAAATCTTTAAAGCATCAAAATTTGTTGTCGC 1002

RESULT 2
AX85550
ID AX85550 standard; cDNA; 1102 BP.
XX
AC AX85550;
XX
DT 07-SEP-1999 (first entry)
XX
DE cDNA of a gene coding for a mouse deleted kin17 protein.
XX
KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW protein interaction; curved DNA; HIV replication; HIV integration;
KW repair enzyme; ss.
XX
OS Mus sp.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PA Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MUF;
PI WPI; 1999-359999/31.
XX
DR New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
PT
PS Claim 4; Page 31; 69pp; French.
XX
CC The present sequence encodes a mouse kin17 protein with amino acids 129-
CC 228 deleted. The mammalian kin17 protein is useful for preparing a
CC medicament for controlling cell proliferation or for controlling
CC fertility. The medicaments can also be used to treat hyperproliferative
CC diseases. Fragments between amino acids 55 and 235 (preferably between
CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
CC regulating the interaction between proteins and curved DNA. The fragment
CC can be used to block replication of HIV or its integration into the human
CC genome or to target repair enzymes to curved DNA sites. Expression
CC vectors for kin17 can be used for controlling cell proliferation
XX
SQ Sequence 1102 BP; 373 A; 205 C; 272 G; 252 T; 0 U; 0 Other;

Query Match 70.3%; Score 704.6; DB 2; Length 1102;
Best Local Similarity 86.1%; Pred. No. 1.6e-183;
Matches 819; Conservative 0; Mismatches 119; Indels 13; Gaps 3;

QY 58 ATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATGCCCAACAGGATCAAGTCCCAAGGG 117
DB 1 ATGGGCAAGTCGGATTTTCTGAGCCCAAGGCCATCGCATAGATTAAGTCCCAAGGG 60

QY 118 CTGCAGAGACTACGCTGGTATTTGCCAGATGTGCCAGAGCAGTGCCTGGACGAGAAATGCG 177
DB 61 CTCCAGAAAGCTTCGCTGGTACTGCCAGATGTGCCAAAAGCAATGCCGCGCAGAAATGCG 120

QY 178 TTTAAGTGTCAATGTATGTCGGAATCTCATCAGAGACAATCTTCTGCTTCAGAAAT 237
DB 121 TTTAAGTGTCAATGTATGTCGGAATCTCATCAAGAGACAATCTTCTGCTTCAGAAAT 180

QY 238 CCTCAGCAGAGTTTATGGATTTATTTTTCAGAGGAATTTCCGAAATGACTTTCTAGAACTTCTC 297
DB 238 CCTCAGCAGAGTTTATGGATTTATTTTTCAGAGGAATTTCCGAAATGACTTTCTAGAACTTCTC 297


```

Db      181 CCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGAAATGACTTTCTGGAACCTTCTG 240
Qy      298 AGGAGAGCGTTTGGCACTAAAGGGTCCACAACAATTGTCTACAACGATACATCAGC 357
Db      241 AGCGAGCGTTTGGCACTAAAGGGTCCACAACAATTGTCTACAATGAATACATCAGC 300
Qy      358 CACCGAGAGCAGATCCACATGAATGCCACTCAGTCAGTGGGMAACTCTGACTGATTTTACTTAAG 417
Db      301 CACCGAGAGCAGATCCACATGAACGTACCAGTGGGAGACATGACCGACTTTTACCAAG 360
Qy      418 TGGCTGGGCAGAGAGGCTTTGTGCAAA-----AGTGCATCAGAGCAGATAGGAAGTTCA 471
Db      361 TGGCTGGGCAGAGAGGCTTTGTGTAAGGTACCAGTGCATGAACTGTCTGGGGAGCGCA 420
Qy      472 GCATCAGTGAACGAAAGAATCTTCCAGAGCTCACTCAGTCACTTAAGAAAGAAAGAGAA 531
Db      421 GCATCCGGGAAACGGAAGAGTCTTTCACAGAGCTCCGCCAGGCT-----GCGAAGAG 474
Qy      532 AAGAAATCTCAGTGGATGAATCATCGAGATTGAAGAGGAAAGAAAGAACTGCCCGA 591
Db      475 AAGAGTCCGCCCTGGATGAGATCATGGAGCTCGAAGAGGAAAGAAAGAGCGACGG 534
Qy      592 ACAGACTACTGGCTACAGCGCTGAAATTTATTTGAAATTTAATACCAAGAAACTGGGAG 651
Db      535 ACAGCGCTGTTTACAGCCGGGATCGTTGTGAAATTTATACGAAGAAGCTTGGGGAG 594
Qy      652 AATATCATAGAAAAAGGCTATTTTAAGAGATTAATTGACAAATATACAGCTGTGTG 711
Db      595 AATATCATCAAGAAAGGGTGGTTAAGGAAGTGAATGACAGGTACACAGCTGTGGTA 654
Qy      712 AAGATGATGATCTGAGACAGCTGAACTGAACTTGACAGACTCAATTTAGAGACAGTAAT 771
Db      655 AAGATGATGATCTGAGACAGCTGAACTGAACTTGACAGACTCAATTTAGAGACAGTAT 714
Qy      772 CCAGCACAGGAAAGAAATTTAGTTTAAATGGAGGCTACAGAGGAATGAAGTACC 831
Db      715 CCGGCCCGGGGAAAGGGTCTAGTTTAAATGGAGGCTACAGAGGAATGAAGCACT 774
Qy      832 CTAGAATCCATCAATGAGAAGACTTTTTCAGCTACTATCGTCAATGAACTGGCCCTTTA 891
Db      775 CTCGAATCCATCAATGAGAGGCTTTTTCAGCCACGATAGTCAATGAACTGGACCTTTG 834
Qy      892 AAAGGACGACAGTTGAGGAATTCATATGAACACATTTCTAACTTGGCTGAGTTGA 951
Db      835 AAAGGACGACAGTTGAGGAATTCATATGAACACATATCTAACTTGGCTGAGTTGA 894
Qy      952 AATTTGTTAACTATACCTTTAAATCTTAAAGCATCAAAATTTGGTTGCG 1002
Db      895 AATTTGTTAACTATACATTTGAATCTGTAACT-GTGAAGCATCAAAATTTGGTTAGC 944

```

RESULT 3

AA85549

ID AA85549 standard; cDNA; 1296 BP.

XX AC

XX AC

XX AC

DT 07-SEP-1999 (first entry)

XX DE

XX DE cDNA of a gene coding for the human kin17 protein.

XX KW Human; kin17 protein; cell proliferation; fertility;

XX KW hyperproliferative disease; protein interaction; curved DNA;

XX KW HIV replication; HIV integration; repair enzyme; ss.

XX OS Homo sapiens.

XX XX

XX PN FR272046-A1.

XX XX

XX PD 11-JUN-1999.

XX XX

XX PF 09-DEC-1997; 97FR-00015536.

XX XX

09-DEC-1997; 97FR-00015536.

(COMS) COMMISSARIAT ENERGIE ATOMIQUE.

Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

WPI; 1999-359999/31.

New DNA coding for human kin17 protein - useful for controlling cell

proliferation or fertility.

Claim 1; Page 30-31; 69pp; French.

The present sequence encodes a human kin17 protein. The mammalian kin17

protein is useful for preparing a medicament for controlling cell

proliferation or for controlling fertility. The medicaments can also be

used to treat hyperproliferative diseases. Fragments between amino acids

55 and 235 (preferably between amino acids 129 and 228) of a mammalian

kin17 protein are useful for regulating the interaction between proteins

and curved DNA. The fragment can be used to block replication of HIV or

its integration into the human genome or to target repair enzymes to

curved DNA sites. Expression vectors for kin17 can be used for

controlling cell proliferation

Sequence 1296 BP; 453 A; 240 C; 310 G; 293 T; 0 U; 0 Other;

Query Match 69.5%; Score 696.4; DB 2; Length 1296;

Best Local Similarity 77.2%; Pred. No. 3.2e-181;

Matches 1001; Conservative 0; Mismatches 1; Indels 294; Gaps 1;

Qy 1 TGATTCCAGCTCGGTACCCGGGATCCGATTAGAAAGTATCGCTGCCGTGGTCCCATG 60

Db 1 TGATTCCAGCTCGGTACCCGGGATCCGATTAGAAAGTATCGCTGCCGTGGTCCCATG 60

Qy 61 GGGAAGTCGATTTTCTTACTCCAGAGCTATGCCAACAGGATCAAGTCCAGGGGCTG 120

Db 61 GGGAAGTCGATTTTCTTACTCCAGAGCTATGCCAACAGGATCAAGTCCAGGGGCTG 120

Qy 121 CAGAGCTACGCTGGTATTCAGAGATGTCAGAGCAGTCCGGGACGAGATGGCTTT 180

Db 121 CAGAGCTACGCTGGTATTCAGAGATGTCAGAGCAGTCCGGGACGAGATGGCTTT 180

Qy 181 AAGTGTCAATGTATGTCGGAATCTCATCAGAGCAACTATTGTGGCTTCAGAAAATCCT 240

Db 181 AAGTGTCAATGTATGTCGGAATCTCATCAGAGCAACTATTGTGGCTTCAGAAAATCCT 240

Qy 241 CAGCAGTTATGATTTATTTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTCAGG 300

Db 241 CAGCAGTTATGATTTATTTTTCAGAGGAATTCGAAATGACTTTCTAGAACTTCTCAGG 300

Qy 301 AGAGCTTTGCACTAAAGGGTCCACAACAATTGCTTACAACGAATACATCAGGCAC 360

Db 301 AGAGCTTTGCACTAAAGGGTCCACAACAATTGCTTACAACGAATACATCAGGCAC 360

Qy 361 CGAGAGCAGATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAGTGG 420

Db 361 CGAGAGCAGATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAGTGG 420

Qy 421 CTGGGACAGAGAGGCTTTGTGCAAA----- 444

Db 421 CTGGGACAGAGAGGCTTTGTGCAAAAGTGGACGAGACACCAGGCTGGTATATTTCAGTAC 480

Qy 445 ----- 444

Db 481 ATAGACAGGACCCAGAAACTATCCCGGGCACTGGAACTGGAGAAAGAAAGAGCAG 540

Qy 445 ----- 444

Db 541 GACCTTGATGATGAAGAAAGAACTGCCAAATTTATTGAAGAGCAAGTGAGAGAGGCTG 600

Qy 445 ----- 444

Db 601 GAAGGAAGGAAAGGAGGCTCCTACTTTTACGGAATTTAAGACAGAGAAATGATGAAGAG 660

Qy	445	-----	444
Db	661	AAAGTCACGTTTAAATTTTGAGTAAAGGAGCATGTAGCTCATCCGAGCAACATCTTCCAAG	720
Qy	445	-----	444
Db	721	TCAAGTACTCTGGACCGAGTGCATGAAGACGATAGAAAGTTTCAGCATCAGTGAACGA	780
Qy	487	AAAGAAATCTTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCAC	546
Db	781	AAAGAAATCTTTCCAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCAC	840
Qy	547	GATGAATCATGGAGATTGAAGAGGAAAGAAAGAACTGCCGAAACAGACTACTGGCTA	606
Db	841	GATGAATCATGGAGATTGAAGAGGAAAGAAAGAACTGCCGAAACAGACTACTGGCTA	900
Qy	607	CAGCCTCAAAATATTGTGAAATTTATAACCAAGAAACTGGGAGAGAAATATCATAAGAAA	666
Db	901	CAGCCTCAAAATATTGTGAAATTTATAACCAAGAACTGGGAGAGAAATATCATAAGAAA	960
Qy	667	AAGCCTATTGTTAAGGAAGTAAATTGA	726
Db	961	AAGCCTATTGTTAAGGAAGTAAATTGA	1020
Qy	727	GGAGACAAGCTGAAACCTTGACCACTCATTTAGACACAGTAAATCCAGCACCAGGAAA	786
Db	1021	GGAGACAAGCTGAAACCTTGACCACTCATTTAGACACAGTAAATCCAGCACCAGGAAA	1080
Qy	787	AGAATCTAGTTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAATCCATCAAT	846
Db	1081	AGAATCTAGTTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCCCTAGAAATCCATCAAT	1140
Qy	847	GAGAAAGACTTTTTTCAGCTACTATCGTCATTTGAAACTGGCCCTTTTAAAGACCGCAGATT	906
Db	1141	GAGAAAGACTTTTTTCAGCTACTATCGTCATTTGAAACTGGCCCTTTTAAAGACCGCAGATT	1200
Qy	907	GAAGGAATTCAAATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAATTTGTTTAAACAT	966
Db	1201	GAAGGAATTCAAATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAATTTGTTTAAACAT	1260
Qy	967	ACCTTTAAAAATCTTAAAGCATCAAAATGGTGTTCGC	1002
Db	1261	ACCTTTAAAAATCTTAAAGCATCAAAATGGTGTTCGC	1296

RESULT 4
AAK85552
ID AAK85552 standard; cDNA; 1002 BP.
XX
XX AC AAX85552;
XX AC
XX DT 07-SEP-1999 (first entry)
XX DE Probe derived from cDNA of a gene coding for the human kinl7 protein.
XX KW Human; kinl7 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme; probe; ss.
XX OS Homo sapiens.
XX PN FR2772046-A1.
XX PD 11-JUN-1999.
XX PF 09-DEC-1997; 97PR-00015536.
XX PR 09-DEC-1997; 97PR-00015536.
XX PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX PI Kannonouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

WIPI; 1999-359999/31.
 New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.
 Claim 7; Page 32-33; 69pp; French.
 The present sequence is derived from nucleotides 207-1208 of AX85549, and is a probe used to isolate human kin17 cDNA. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation
 Sequence 1002 BP; 374 A; 179 C; 235 G; 214 T; 0 U; 0 Other;
 Query Match 46.9%; Score 470; DB 2; Length 1002;
 Best Local Similarity 100.0%; Pred. No. 7.5e-119;
 Matches 470; Conservative 0; Mismatches 0; Indels 0; Gaps 0
 445 AGTGCACCTGAAGACGATAGGAAGTTTCAGCATCAGTGAACGAAAAAGAAATCTTCCAGAGC 504
 533 AGTGCACCTGAAGACGATAGGAAGTTTCAGCATCAGTGAACGAAAAAGAAATCTTCCAGAGC 592
 505 TCAACTCAGTCTAAAGAAAAAGAAAGAAATCTGCATGGATGAAATCATGAGATT 564
 593 TCAACTCAGTCTAAAGAAAAAGAAAGAAATCTGCATGGATGAAATCATGAGATT 652
 565 GAAGAGGAAAGAAAGAAAGAAATCTGCCGACAGACTACTGGCTACAGCCTCAATATTATTGTG 624
 653 GAAGAGGAAAGAAAGAAAGAAATCTGCCGACAGACTACTGGCTACAGCCTCAATATTATTGTG 712
 625 AAAATTATTAACCAAGAAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAA 684
 713 AAAATTATTAACCAAGAAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAA 772
 685 GTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAGCTGAAACTT 744
 773 GTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAGCTGAAACTT 832
 745 GACCAGACTCAATTTAGACAGAGTAATTCACAGACAGGAAAGAAAGAAATCTAGTGTTTAAAT 804
 833 GACCAGACTCAATTTAGACAGAGTAATTCACAGACAGGAAAGAAAGAAATCTAGTGTTTAAAT 892
 805 GGAGGCTACAGAGGAAATGAAGGTACCTTAGAATCCATCAATGAGAAGACTTTTTTCAGCT 864
 893 GGAGGCTACAGAGGAAATGAAGGTACCTTAGAATCCATCAATGAGAAGACTTTTTTCAGCT 952
 865 ACTATCGTCAATTGAACCTGGCCCTTTTAAAGGACGACAGATTGAAGGAAT 914
 953 ACTATCGTCAATTGAACCTGGCCCTTTTAAAGGACGACAGATTGAAGGAAT 1002
 RESULT 5
 ADQ56943
 ID ADQ56943 standard; DNA; 679 BP.
 XX AC ADQ56943;
 XX
 DT 21-OCT-2004 (first entry)
 XX
 DE Novel canine microarray-related DNA sequence SeqID8245.
 XX
 KW canine microarray; drug screening; toxicity assay;
 KW environmental pollutant; cellular response; gene expression profile;
 KW toxic response; liver necrosis; fatty liver disease;
 KW protein adдук formation; hepatitis; dog; da.

Best Local Similarity 66.1%; Pred. No. 4.7e-104;		
Matches 819; Conservative 0; Mismatches 119; Indels 301; Gaps 3;		
QY	58	ATGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAACAGGATCAAGTCCAAGGGG 117
Db	1	ATGGGCAAGTCGGATTTTCTGAGCCCAAGGCCATCGCCAATAGAAATTAAGTCCAAGGG 60
QY	118	CTCAGAAGCTACGCTGGTATTCAGATGTGCCAAGCAGTGC CGGACAGAAATGGC 177
Db	61	CTCAGAAGCTTCGGTGGTACTGCCAGATGTGCCAAAGCAATGCCCGACGAGAAATGGC 120
QY	178	TTTAACTGTCTATGTATGTCGGAATCTCATCAGAGACAATATTGCTGCTTCAGAAAT 237
Db	121	TTTAACTGTCTATGTATGTCGGAATCTCATCAAGACAACCTGTTGCTGGCTTCAGAAAC 180
QY	238	CCTCAGCAGTTTATGGATTTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTC 297
Db	181	CCTCAGCAGTTTATGGATTTATTTTTCAGAGGAATTCGGAATGACTTTCTGGAATCTCTG 240
QY	298	AGGAGACGCTTTGGCACTAAAGGTTCCACAACAACTTGTCTAACCAATACATCAGC 357
Db	241	AGGCGACGCTTTGGCACTAAAGGTTCCACAACAACTTGTCTACAATGAATACATCAGC 300
QY	358	CACCGAGACACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAG 417
Db	301	CACCGAGACACATCCACATGAACGCTACCCAGTGGGAGACACTGACCGACTTTTACCAAG 360
QY	418	TGGCTGGGCAGAGAAGGCTTTGTGCAA----- 443
Db	361	TGGCTGGGCAGAGAGGCTTTGTGTAAGTGGATGAGACACCGAAAGGCTGGTACATTCAG 420
QY	444	----- 443
Db	421	TACATAGACAGAGACCCAGAAACCATCCGTCGGCAACTGGAAATTAGAAAAAAGAGAAG 480
QY	444	----- 443
Db	481	CAAGATCTGGACGATGAAGAAAAAATCGCAAGTTCATTGAGGAGCAGGTGAGAAGAGGC 540
QY	444	----- 443
Db	541	CTGGAAGGGAAGAGCAGGAGAGACACCTGTTTTTACAGAACTTAGCCGAGAAATGAGGAA 600
QY	444	----- 443
Db	601	GAAGAAAGTTACGTTCAATCTGAATAAAGGAGCGGCTGGCTCAGCGGGAGCTACAACATCC 660
QY	444	----- AAGTGCACCTGAAGACGATAGGAAGTTTCAGCATCAGTGAAG 483
Db	661	AAGTCAAGCTCTTTGGGACCAAGTGCACTGAAGCTCTCTGGGAGCGCAGCATCCGGGAAA 720
QY	484	CGAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAGAAAGAAAGAAATCTGCA 543
Db	721	CGAAAGAGTCTTTCACAGAGCTCCGCCAGCT-----GCGAAGAAAGAAAGTTCGCC 774
QY	544	CTGGATGAATCATGAGATTTGAAGAGAAAAAGAAAGAAAGAAAGAAAGAAATCTATCG 603
Db	775	CTGGATGAGATCATGAGCTCGAAGAGGAAAAAGAAAGGACCGCACGGACAGACGCTCG 834
QY	604	CTACAGCCTGAATTAATTTGTGAATAATTATAACCAAGAAACTGGGAGAGAAATATCATTAAG 663
Db	835	TTACAGCCGGGATCGTTGTGMAAATTTATAACGAAGAAAGCTTTGGGAGAAATATCACAA 894
QY	664	AAAAAGCTATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGAT 723
Db	895	AAGAAAGGGGTCGTTAAGGAGTATTGACAGGTACACAGCTGTGGTAAAGATGACTGAC 954
QY	724	TCTGGAGACAGCTGAAACTTGACAGACTCATTTAGAGACAGTAATTTCCAGCACGAGGA 783
Db	955	TCTGGAGACAGGCTGAAACTTGGACACAGACTCATTTTAGNACAGTCAATTCGGCCCGGG 1014
QY	784	AAAAAGAAATCTAGTTTTAAATGGAGGCTACAGAGGAAATGAAGGTACCTTAGAATCCATC 843

Db	1015	AAAAGGGTCTAGTTTTTAATGGAGGCTACAGAGAAATGAAGGCACCTCTCGAATCCATC 1074	
QY	844	AATGAGAAGACTTTTTCAGCTACTATCGTCAATGAACTGGCCCTTTAAAAGGACGCAGA 903	
Db	1075	AATGAGAAGGCTTTTTCAGCCACGATAGTCAATGAACTGGACCTTTGAAAGGACGCAGA 1134	
QY	904	GTTGAAGGAATTCATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAAAATTTGTTAAC 963	
Db	1135	GTTGAAGGATTTCAATATGAAGACATATCTAACTTGTGAGTTTGAAAAATTTGATAAC 1194	
QY	964	AATACCTTTAAATCTTAAAGCATCAAAATTTGGTGTTCG 1002	
Db	1195	AACACATTGAACT-GTGAAGCATCAAAATTTGTTAGC 1232	
RESULT 7			
AAQ79936			
ID	AAQ79936	standard; cDNA; 1458 BP.	
XX	AAQ79936;		
XX	XX		
DT	25-MAR-2003 (revised)		
DT	06-SEP-1995 (first entry)		
XX	XX		
DE	Murine Kin17 cDNA.		
XX	XX		
KW	chromosomal rearrangement; kin17 protein; SOS DNA repair system; RecA; genotoxic agent; zinc finger; DNA binding protein; ss.		
OS	Mus musculus.		
XX	XX		
FT	Key	Location/Qualifiers	
FT	primer_bind	complement(1..21)	
FT		/*tag= b	
FT		/note= "Oligo L (AAQ79946) binds to complementary strand"	
FT	misc_feature	22..1434	
FT		/*tag= l	
FT		/label= kin17_cDNA	
FT		/note= "nucleotides 1-1414; the genomic DNA contains at least two introns within this sequence, see Comments"	
FT	primer_bind	32..49	
FT		/*tag= c	
FT		/note= "Oligo C (AAQ79938) binding site"	
FT	CDS	46..1221	
FT		/*tag= a	
FT		/product= "Kin17"	
FT		/note= "N's in the sequence denote illegible residues"	
FT	primer_bind	complement(67..86)	
FT		/*tag= d	
FT		/note= "Oligo S (AAQ79947) binds to complementary strand"	
FT	primer_bind	274..297	
FT		/*tag= e	
FT		/note= "Oligo D (AAQ79939) binding site"	
FT	primer_bind	complement(339..360)	
FT		/*tag= f	
FT		/note= "Oligo K (AAQ79945) binds to complementary strand"	
FT	primer_bind	451..474	
FT		/*tag= g	
FT		/note= "Oligo J (AAQ79944) binding site"	
FT	primer_bind	complement(550..567)	
FT		/*tag= h	
FT		/note= "Oligo E (AAQ79940) binds to complementary strand"	
FT	primer_bind	802..825	
FT		/*tag= i	
FT		/note= "Oligo F (AAQ79941) binding site"	
FT	primer_bind	complement(839..862)	
FT		/*tag= j	
FT		/note= "Oligo G (AAQ79942) binds to complementary strand"	
FT	primer_bind	complement(1435..1458)	
FT		/*tag= k	
FT		/note= "Oligo B (AAQ79937) binds to complementary strand"	
XX	XX		
PN	FR2706487-Al.		

XX 23-DEC-1994.
XX PD
XX PF 15-JUN-1993; 93FR-00007171.
XX PR 15-JUN-1993; 93FR-00007171.
XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX PA Angulo-Mora JF, Tissier A, Frelat G, Mauffrey P, Guilly M;
XX PI WPI; 1995-039031/06.
XX DR Purified murine kin17 protein prepn. for detecting chromosomal
XX PT rearrangements - also related antibodies, human and murine DNA, primers,
XX PT probes and vectors, used to assess damage caused by genotoxic agents.
XX PS Claim 9; Page 33; 54pb; French.
XX CC The murine kin17 protein includes a zinc finger domain (see AAR66766),
CC recognises single- and double-stranded DNA (partic. regions of secondary
CC structure), has apparent mol. wt. 43 kD and is recognised by both anti-
CC kin17 antibodies and antibodies against the RecA protein of E.coli. The
CC kin17 protein is involved in DNA repair; it can be used to monitor
CC chromosomal rearrangements following exposure to genotoxic agents. The
CC kin17 cDNA sequence A079936 consists of a 1414 nucleotide sequence,
CC flanked by primer binding sites; the genomic kin17 DNA is claimed in
CC which an intron is inserted at position 137 of the 1414 nucleotide cDNA
CC and a second intron is located between nucleotides 339-429 of the cDNA.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 1458 BP; 496 A; 272 C; 372 G; 311 T; 0 U; 7 Other;
Query Match 41.4%; Score 415.2; DB 2; Length 1458;
Best Local Similarity 65.2%; Pred. No. 1.2e-103;
Matches 826; Conservative 0; Mismatches 139; Indels 301; Gaps 3;
31 TAGAAGTATGCTGCTGCGTGTGCGCATCGGAGTCGGATTTCTTACTCCCAAGGCT 90
19 TCGGTAGTTGAGCGCTCAGGGTTCGCGATGGGAATCGGATTTCTGAGCCCCAAGGCC 78
91 ATGCCAACAGGATCAAGTCCAGGGGCTCGAGAGCTACGCTGGTATTCGCCAGATGTC 150
79 ATGCCCAATAGATTAAGTCCAAAGGCTCCAGAACGTTNNNTACTCCAGATGTC 138
151 CAGAAGCAGTGCCTGCGGACGAGAAATGGCTTTAAGTGTCAATTTGATGTCGGAATCTCATCAG 210
139 CAAAAGCAATGCTCGCGACGAGAAATGGCTTTAAGTGTCACTGTATGTCGAATCTCATCAA 198
211 AGACAATATGCTGGCTTCAGAAATCCTCAGCAGTTTATGATTTATTTTCAGAGGAA 270
199 AGACAATGTTGCTGGCTTCAGAAACCCCTCAGCAGTTTATGGAATTTATTTTCAGAGGAA 258
271 TTCGAAATGACTTTCTAGAACTTCTCAGAGACGCTTTTGGCACTAAAGGGTCCACAAC 330
259 TTCGAAATGACTTTCTGGAATCTCTGAGGCGAGCTTTTGGCACTAAAGGGTCCACAAC 318
331 AACATTGCTTACACGAATACATCAGCCACCGAGACACATCCACATGAATGCCACTCAG 390
319 AACATTGCTTACATGAATACATCAGCCACCGAGACACATCCACATGAACGCTTACCAG 378
391 TGGGAAACTCTGACTGATTTTACTAAGTGGCTGGGAGAGAGGCTTTGTGCA----- 443
379 TGGGAGACACTGACCGACTTTTACCAGTGGCTGGGAGAGAGGCTTTGTGTAAGTGGAT 438
444 ----- 443
439 GAGACCGAAAGGCTGTATCATTTTCACTAGTACATAGACAGAGCCAGAAACCATCCGTCGG 498
444 ----- 443
499 CAACCTGGAAATTAGAAAAAAGAGAACAGATCTGGACGATGAAGAAAAAACTGCCAAG 558

QY 444 ----- 443
Db 559 TTCAATTGAGGAGCAGGTCAAGAGAGGCTTNGAAGGAAAGAGCAGGAGACACCTGTTTTT 618
QY 444 ----- 443
Db 619 ACAGAACTTTAGCCGAGAAAAATGAGGAAGAAAGTTACGTTCAATCTGAATTAAGGAGCG 678
QY 444 ----- 456
Db 679 GGTGCTCAGCGGAGCTTACAACATCCRAAGTCAAGCTCTTTGGAGCAAGTGCACCTGAAG 738
QY 457 ACGATAGGAAAGTTTACAGCATCAGTGAAGAAAGAAAGTCTTCCAGAGCTCAACTCAGTCT 516
Db 739 CTGCTGGGAGCGCAGCATCCGGAACCGAAAGAGTCTTTCACAGAGCTCCGCCAGCCT 798
QY 517 AAGAAAGAAAGAAAGAAATCTGCACCTGGATGAATCATCGAGATTGGAAGAGAAAG 576
Db 799 -----GCCAAGAAAGAAAGTCCGCCCTGGATGAGATCATGGAGCTCGAAGAGGAAAG 852
QY 577 AAAAGAACTGCCCGAACAGACTTACTGGCTACAGCCTCGAAATTTATTTGTAATAATTATAACC 636
Db 853 AAAAGAACCGCAGGACAGCAGCCTGTTTACAGCCGGGATCGTTGTGAAATTTATAACG 912
QY 637 AAGAAACTCGGAGAGAAATATATCAAGAAAGAGCTTATTTGTAAGGAAGTAAATTGACAAA 696
Db 913 AAGNAGCTTTGGGAGAAATATCAAGAAAGAAAGGGTCTGTTAAGGAGTGAATTGACAGG 972
QY 697 TATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAGCTGAACTTGCACAGACTCAT 756
Db 973 TACACAGCTGTGTAAAGATGACTGACTCTGGAGACAGGCTGAACTGGACCCAGACTCAT 1032
QY 757 TTAGAGACAGTAAATCCAGCACCCAGGAAAGAAATTTCTAGTTTAAATGGAGGCTACAGA 816
Db 1033 TTAGAGACAGTAAATCCGCGCCCGGGAAGAAAGGTTCTAGTTTAAATGGAGGCTACAGA 1092
QY 817 GGAATTAAGAGGTACCCCTAGAAATCCATCAATGAGAAAGCTTTTTCAGCTACTATCGTCATT 876
Db 1093 GGAATTAAGAGGCACTCTCGAAATCCATCAATGAGAGGCTTTTTCAGCCACGATAGTCATT 1152
QY 877 GAACTGGCCCTTTAAAGAGCCAGACTTGAAGGAATTCATATATGAAGCATTTCTTAA 936
Db 1153 GAACTGGACCTTTGAAGAGCCAGAGTTGAAGGTATTTCAATATGAAGCATATCTTAA 1212
QY 937 CTGCTGCTGAGTTTGAAGAAATTTGTTAAACAATACCTTTAAATCTTAAAGCATCAAATGGT 996
Db 1213 CTGCTGCTGAGTTTGAAGAAATTTGATAACAACATTCGAACT-CTGAAGCATCAAATGGT 1271
QY 997 GTTCGC 1002
Db 1272 GTTAGC 1277
RESULT 8
AAH34095
ID AAH34095 standard; cDNA; 591 BP.
XX AAH34095;
AC AC
XX XX
DT 03-SEP-2001 (first entry)
XX Human colon cancer antigen encoding cDNA SEQ ID NO:1177.
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma; ss.
XX Homo sapiens.
XX OS
XX WO200122920-A2.
XX PN
XX 05-APR-2001.
XX PD
XX 28-SEP-2000; 2000WO-US026524.
PF

XX 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI; 2001-235357/24.
XX P-PSDB; AAG74690.
DR Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 1; Page 2977-2978; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
XX Sequence 591 BP; 199 A; 95 C; 114 G; 182 T; 0 U; 1 Other;
SQ
Query Match 26.7%; Score 267.2; DB 4; Length 591;
Best Local Similarity 95.3%; Pred. No. 4.9e-63;
Matches 286; Conservative 0; Mismatches 13; Indels 1; Gaps 1;
QY 703 GCTGTTGTGAAGATGATGCTGAGACAAGCTGAAACTTGACCAGACTCATTTAGAG 762
DB 28 GCTGCAGGAATTCGGCAGCAGCTGGAGACAAGCTGAAACTTGACCAGACTCATTTAGAG 87
QY 763 ACAGTAATTCACGACCAGGAAAAAGAAATCTAGTTTTAAATGGAGGCTACAGAGGAAAT 822
DB 88 ACAGTAATTCACGACCAGGAAAAAGAAATCTAGTTTTAAATGGAGGCTACAGAGGAAAT 147
QY 823 GAAGTACCTTGAATCCATCAATGAGAAGACTTTTTCAGCTACTATCGTCAATTGAAACT 882
DB 148 GAAGTACCTTGAATCCATCAATGAGAAGACTTTTTCAGCTACTATCGTCAATTGAAACT 207
QY 883 GGCCTTTTAAAGGACGAGTTGAAGAAATTCATATGAAGACTTTCTAAACTTGGC 942
DB 208 GGCCTTTTAAAGGACGAGTTGAAGAAATTCATATGAAGACTTTCTAAACTTGGC 267
QY 943 TCAGTTTGAATAATTTGTTAAACATACCTTTTAAATCTTAAAGCATCAAAATTTGGTGTTCG 1002
DB 268 TCAGTTTGAATAATTTGTTAAACATA-CATTAAATCTTAAAGCATCAAAATTTGGTGTTCG 326
RESULT 9
ADP22451
ID ADP22451 standard; DNA; 1372 BP.
XX
XX AC ADP22451;
XX
XX 12-AUG-2004 (first entry)
XX
XX DE Sea-squirt (Ciona intestinalis) zinc finger protein coding sequence #13.
XX
XX sea-squirt; zinc finger protein; gene detection; drug development;

zinc finger protein-associated disease; gene; ds.
Ciona intestinalis.
JP2004057126-A.
26-FEB-2004.
31-JUL-2002; 2002JP-00222484.
31-JUL-2002; 2002JP-00222484.
(KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
WPI; 2004-208711/20.
P-PSDB; ADP22452.
Novel gene encoding zinc finger protein, useful as probe in gene
detecting instruments and in development of drug for treating zinc finger
protein associated diseases.
Claim 3; SEQ ID NO 25; 972bp; Japanese.
The invention comprises the amino acid and coding sequences of sea-squirt
(Ciona intestinalis) zinc finger proteins. The DNA and protein sequences
of the invention are useful in a gene detecting instrument. The DNA and
protein sequences of the invention are useful in the development of drugs
for the treatment of zinc finger protein-associated diseases. The present
DNA sequence encodes a sea-squirt zinc finger protein of the invention.
Sequence 1372 BP; 500 A; 224 C; 286 G; 362 T; 0 U; 0 Other;
Query Match 20.1%; Score 201.4; DB 12; Length 1372;
Best Local Similarity 63.1%; Pred. No. 1e-44;
Matches 310; Conservative 0; Mismatches 181; Indels 0; Gaps 0;
QY 61 GGGAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAACAGATCAAGTCCAAGGGCTG 120
DB 52 GAGAAACAGGATTTCTCACTGCCAAGGCTATCGGCAACCGTATCAATCAANGGACTT 111
QY 121 CAGAAAGTACGCTGTTATGGCAGATGTGCCAGAGCAGTGCCTGGGACGAGATGGCTTT 180
DB 112 CAAAAGTTCGATGTTATGTCAAATGTGCCAAAACAGTGCAGAGATGAGAACGGTTT 171
QY 181 AAGTCTATTGATGTCGGAATCTCATCAGAGACAATCTATTGCTGGCTTCAGAAAATCCT 240
DB 172 AAATGCCATTGATGTCGGAATCATCATGCGCCAATTTGTTAAATGCTGGTGAATACT 231
QY 241 CAGCAGTTTATGATTTATTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTCTCAGG 300
DB 232 GGAGCTTTTCAGATTTCTTTTCAAGTTCCTTTTCAAGACTTTTATGCAACTTTTGAA 291
QY 301 AGACCTTTTGGCCTTAAAGGGTCCCAACAACATTTGTCTACAACGAATACATCAGCCAC 360
DB 292 ACACCTTTTCGGACTTAAAGAGTTTCATAACAATATTGTATACGAGTATATTTCCTCAT 351
QY 361 CGAGAGCATCCCATGAAATGCCACTCAGTGGGAACTCTGACTGATTTTCTACTAAGTGG 420
DB 352 AAAGAACACATTCATATGAACGCTACAAAATGGGTGACGTTAACTGCTTCAAAAATGG 411
QY 421 CTGGGACAGAGAGGCTTCTGCAAAAGTGCACATGAAGACGATAGGAAGTTTCAGCATCAGT 480
DB 412 TTGGGTAGAGAGGGACACTGTAAAGTGGATCAACCGAAAAGGGTGTGTTTCATCATATAC 471
QY 481 AAACGAAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAAGAAAAGAAATCT 540
DB 472 ATTGATAGAGACCCATTGCTGCTGCCACAGACAGAAAGAACTTGTATAAACAAGCAAG 531
QY 541 GCACCTGGATGA 551
DB 532 GAGAGAGATGA 542

RESULT 10

ADT17629
ID ADT17629 standard; cDNA; 1713 BP.

XX AC
XX ADT17629;

XX DT 13-JAN-2005 (first entry)

XX DE Plant cDNA, Seq ID 2955.

XX KW Plant; ss; gene; transgenic; cold tolerance; growth rate;
XX KW drought tolerance; disease resistance; galactomannan production;
XX KW plant growth regulator; heat tolerance; herbicide tolerance;
XX KW lignin production; extreme osmotic condition tolerance;
XX KW pathogens resistance; pest resistance; yield improvement; seed oil yield;
XX KW seed protein yield.

XX OS Viridiplantae.

XX PN US2004216190-A1.

XX PD 28-OCT-2004.

XX PF 18-DEC-2003; 2003US-00739930.

XX PR 28-APR-2003; 2003US-00424599.

XX PR 28-APR-2003; 2003US-00425115.

XX PA (KOVA/) KOVALIC D K.

XX PI Kovalic DK;

XX DR WPI; 2004-757369/74.

XX PT New recombinant DNA constructs useful in the field of biochemistry and
XX PT genetics, and in particular for producing transgenic plants with improved
XX PT biological characteristics.

XX PS Claim 1; SEQ ID NO 2955; 14pp; English.

XX CC The invention relates a recombinant DNA construct comprising a
XX CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:
XX CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences
XX CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,
XX CC Arabidopsis, wheat and rape but the specification does not indicate which
XX CC sequences is derived from which organism. Also included is a method of
XX CC producing a plant having an improved property, comprising transforming a
XX CC plant with a recombinant DNA construct comprising a promoter region
XX CC functional in a plant cell operably joined to a polynucleotide encoding a
XX CC polypeptide associated with the property, and growing the transformed
XX CC plant. The property is selected from improving plant cold tolerance, for
XX CC manipulating growth rate in plant cells by modification of the cell cycle
XX CC pathway, for improving plant drought tolerance, for providing increased
XX CC resistance to plant disease, for galactomannan production, for production
XX CC of plant growth regulators, for improving plant heat tolerance, for
XX CC improving plant tolerance to herbicides, for increasing the rate of
XX CC homologous recombination in plants, for lignin production, for improving
XX CC plant tolerance to extreme osmotic conditions, for improving plant
XX CC tolerance to pathogens or pests, for yield improvement by modification of
XX CC photosynthesis, for modifying seed oil yield and/or content, for
XX CC modifying seed protein yield and/or content, for yield improvement by
XX CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake
XX CC and for yield improvement by providing improved plant growth and
XX CC development under at least one stress condition. The polynucleotide may
XX CC also encode a plant transcription factor. The methods and compositions of
XX CC the present invention are useful in the field of biochemistry and
XX CC genetics, in particular for producing transgenic plants with improved
XX CC biological characteristics such as increased yield, improved nitrogen
XX CC flow, increasing plant tolerance to cold or heat, improving plant
XX CC tolerance to extreme osmotic and drought conditions, and improving plant
XX CC tolerance to plant pests or pathogens. They can also be used in physical
XX CC arrays of molecules, plant breeding markers, computer-based storage and
XX CC analysis systems. The present sequence is one of the 5544 plant cDNA

CC sequences of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.

XX SQ Sequence 1713 BP; 412 A; 406 C; 522 G; 373 T; 0 U; 0 Other;

Query Match 17.6%; Score 176; DB 13; Length 1713;
Best Local Similarity 63.9%; Pred. No. 1.1e-37;
Matches 266; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY	54	CGCCATGGGGAAGTCGGATTCTTACTCCCAAGGCTATCGCAACAGATCAAGTCCAA	113
DB	171	CGCGATGGGGAAGCACGAGTTCTTACGCGGAAGCGATCGCAACCCGATCAAGCGAA	230
QY	114	GGGGCTGCAGAGTACGCTGTTATTCAGAGATGTCAGAGAGCAGTCCGGGACGAGAA	173
DB	231	GGGGCTGCAGAGTACGCTGTTATTCAGAGATGTCAGAGAGCAGTCCGGGACGAGAA	290
QY	174	TGGCTTTAAGTGTCAATTGTATGTCGGAATCTCATCAGAGACAACTATTGCTGGCTTCA	233
DB	291	CGGGTTCAAGTGCCTACTGCTATGTCGAGTCCGACCGGCGAGATGCAGGTGTTCCGCAT	350
QY	234	AAATCCTCAGCAGTTATGGAATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACT	293
DB	351	GGCGCCGACCGCGTCTCGAGGGCTTCTCCGAGGAATTCCTCGAGTCTCTCTCCCT	410
QY	294	TCTCAGGAGACGCTTTGGCACTAAAGGGTCCACACACATTTGCTACACGAATACAT	353
DB	411	CATCCGCCCGCGCACCGCCACTCCCGCGTCCGCGCCACCGTTGTTACACAGATACAT	470
QY	354	CAGCCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTAC	413
DB	471	CGCGGACCGGACCGACCTCCATGAATCTACGCGGTGGCCACGCTCACCAGTTCGT	530
QY	414	TAAAGTGGCTGGGACAGAAAGGCTTGTGCAAAAGTGCCTCAGAGACATAGGAAGTT	469
DB	531	CAAGCTCTGGGCGCGAAGGGTACTGCAAGGTTGAGAGACACGCCCAAGGGTGGT	586

RESULT 11

AAA52591

ID AAA52591 standard; cDNA; 1780 BP.

XX AC AAA52591;

XX DT 27-SEP-2000 (first entry)

XX DE cDNA encoding maize KIN17 orthologue, ZmKINH-3.

XX KW ZmKINH-3; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
XX KW nuclear localisation; nonhomologous recombination;
XX KW illegitimate recombination; double stranded DNA binding; curved DNA;
XX KW homologous gene targeting; transgenic plant; ss.

XX OS Zea mays.

XX FH Key Location/Qualifiers
XX CDS 172..1443

XX FT /*tag= a
XX FT /product= "Maize KIN17 orthologue, ZmKINH-3"
XX FT /function= "Binds double-stranded curved DNA"

XX PN WO200024900-A1.

XX PD 04-MAY-2000.

XX PF 06-OCT-1999; 99WO-US023280.

XX PR 27-OCT-1998; 98US-0105802P.

XX PA (PTON-) PIONEER HI-BRED INT INC.

XX XX

```
PI Mahajan PB;
XX WPI; 2000-350741/30.
XX Nucleic acids encoding maize KIN17 orthologue proteins useful for
PT preventing illegitimate recombination in cells.
XX Claim 1; Page 77-79; 84pp; English.
XX This sequence represents cDNA encoding the maize KIN17 orthologue ZmKINH-
CC 3. The invention relates to maize KIN17 orthologues (AAB03063- AAB03065),
CC nucleic acids encoding them (AAAS2589-A52591), and expression vectors,
CC transgenic plants and plant seeds comprising nucleotides encoding maize
CC KIN17 orthologues. KIN17 has, until now, been found only in animal
CC (avian, rodent and human) cells, this invention being the first report
CC describing the presence of KIN17 in plants. Murine KIN17 was found to
CC have significant homology to Escherichia coli RecA protein, and contains
CC a zinc finger motif and a nuclear localisation signal. KIN17 binds double
CC -stranded DNA, preferentially binding to curved DNA, and forms
CC intranuclear foci on overexpression in mammalian cells. It is also
CC induced on exposure to gamma or ultraviolet radiation. These findings
CC indicate that KIN17 plays a role in non-homologous (illegitimate)
CC recombination, which occurs at higher rates among higher eukaryotes,
CC particularly plants. Illegitimate recombination in plants is a major
CC impediment to the generation of transgenic crops such as maize. Maize
CC KIN17 orthologue nucleic acid sequences may be used to generate
CC transgenic plants. The transgenic plants generated can be monocots or
CC dicots and are particularly maize, soybean, sunflower, sorghum, canola,
CC wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
CC sequences may be used to reduce KIN17 levels in embryogenic callus or
CC embryo cells, thereby reducing the amount of non-homologous recombination
CC and enhancing homologous gene targeting
XX SQ Sequence 1780 BP; 467 A; 408 C; 525 G; 380 T; 0 U; 0 Other;
Query Match 17.6%; Score 176; DB 3; Length 1780;
Best Local Similarity 63.9%; Pred. No. 1.2e-37;
Matches 266; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 54 CGCCATGGGGAAAGTCGGATTTCTTACTCCCAAGGCTATCGCAACAGCATCAAGTCCAA 113
DB 168 CGCGATGGGGAGACGACGAGTTCTGACGCGGAGGCGATCGGAAACGGATCAAGCGAA 227
QY 114 GGGGCTGCAGAACCTAGCTGGTATTGCGCAGATGTCGCAAGACGATCGCGGACGAGAA 173
DB 228 GGGGCTGCAGAACCTAGCTGGTATTGCGCAGATGTCGCAAGACGATCGCGGACGAGAA 287
QY 174 TGGCTTTAAGTGTCATTGTATGTCGCAATCTCATCAGACAACTATTGCTGCTTCAGA 233
DB 288 CGGGTTCAAGTGCACCTGCATGTCGAGTTCGACACGCGCAGATCGAGTGTTCGGCAT 347
QY 234 AAATCCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACT 293
DB 348 GGGCCCGACCGCGTCTGTCAGGGCTTCTCCGAGGAATTCCTCGAGTCTCTCTCCCT 407
QY 294 TCTCAGAGACGCTTTGGCAGCTAAAGGGTCCACACACATTTGCTTACACGAATACAT 353
DB 408 CATCCGCGCGCGCACCGCCACTCCCGCGCTGCGCGCACCGTTGTCTACAAACAGTACAT 467
QY 354 CAGCCACCGAGACACATCCACATGAATGCACCTCAGTGGAACTCTGACTGATTTTAC 413
DB 468 CGCGGACCGGCACACGTCACATGAATCTTACCGGTTGGCCACCGCTCACCAGTTCGT 527
QY 414 TAAATGGCTGGGAGAGAGGCTTTGTGCAAAAGTGCACCTGAAGACGATAGGAGTT 469
DB 528 CAAAGTCTCTGGGCGCGGAAGGGTACTGCAAGGTTGAGGACACGCCCAAGGGGTGT 583
RESULT 12
ID AAAS2590
XX AAAS2590 standard; cDNA; 1780 BP.
XX AC AAAS2590;
```

```
XX 27-SEP-2000 (first entry)
XX cDNA encoding maize KIN17 orthologue, ZmKINH-2.
XX ZmKINH-2; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
KW nuclear localisation; nonhomologous recombination;
KW illegitimate recombination; double stranded DNA binding; curved DNA;
KW homologous gene targeting; transgenic plant; ss.
XX Zea mays.
XX Key Location/Qualifiers
CDS 172..1443
FT /tag= a
FT /product= "Maize KIN17 orthologue, ZmKINH-2"
FT /function= "Binds double-stranded curved DNA"
XX WO200024900-A1.
XX 04-MAY-2000.
XX 06-OCT-1999; 99WO-US023280.
XX 27-OCT-1998; 98US-0105802P.
XX (PION-) PIONEER HI-BRED INT INC.
XX Mahajan PB;
XX WPI; 2000-350741/30.
XX P-PSDB; AAB03064.
XX Nucleic acids encoding maize KIN17 orthologue proteins useful for
PT preventing illegitimate recombination in cells.
XX Claim 1; Page 74-76; 84pp; English.
XX This sequence represents cDNA encoding the maize KIN17 orthologue ZmKINH-
CC 2. The invention relates to maize KIN17 orthologues (AAB03063- AAB03065),
CC nucleic acids encoding them (AAAS2589-A52591), and expression vectors,
CC transgenic plants and plant seeds comprising nucleotides encoding maize
CC KIN17 orthologues. KIN17 has, until now, been found only in animal
CC (avian, rodent and human) cells, this invention being the first report
CC describing the presence of KIN17 in plants. Murine KIN17 was found to
CC have significant homology to Escherichia coli RecA protein, and contains
CC a zinc finger motif and a nuclear localisation signal. KIN17 binds double
CC -stranded DNA, preferentially binding to curved DNA, and forms
CC intranuclear foci on overexpression in mammalian cells. It is also
CC induced on exposure to gamma or ultraviolet radiation. These findings
CC indicate that KIN17 plays a role in non-homologous (illegitimate)
CC recombination, which occurs at higher rates among higher eukaryotes,
CC particularly plants. Illegitimate recombination in plants is a major
CC impediment to the generation of transgenic crops such as maize. Maize
CC KIN17 orthologue nucleic acid sequences may be used to generate
CC transgenic plants. The transgenic plants generated can be monocots or
CC dicots and are particularly maize, soybean, sunflower, sorghum, canola,
CC wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
CC sequences may be used to reduce KIN17 levels in embryogenic callus or
CC embryo cells, thereby reducing the amount of non-homologous recombination
CC and enhancing homologous gene targeting
XX SQ Sequence 1780 BP; 467 A; 408 C; 525 G; 380 T; 0 U; 0 Other;
Query Match 17.6%; Score 176; DB 3; Length 1780;
Best Local Similarity 63.9%; Pred. No. 1.2e-37;
Matches 266; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 54 CGCCATGGGGAAAGTCGGATTTCTTACTCCCAAGGCTATCGCAACAGCATCAAGTCCAA 113
DB 168 CGCGATGGGGAGACGACGAGTTCTTACCGCGGAGGCGATCGGAAACGGATCAAGCGAA 227
QY 114 GGGGCTGCAGAACCTAGCTGGTATTGCGCAGATGTCGCAAGACGATCGCGGACGAGAA 173
```


Db 228 GGGGCTGCAGAGCTGGGTGGTACTGCCAGATGTGTCAAGACAGTGGCGACAGAA 287
 QY 174 TGGCTTTAAGTGTCAATGTATGTCGGAATCTCATCAGAGACAATATTTGCTGCTTCA 233
 Db 288 CGGGTTCAAGTGCACCTGCTATGTCGAGTGGCAGCGGAGCAGATGAGGTTCGGCAT 347
 QY 234 AATCTCTCAGCAGTTATGATTTATTTTCAGAGGAATTCGGAATGACTTTCTAGACT 293
 Db 348 GGGCGCGGACCGGTGCTGAGGGCTTCTCGAGGAATTCCTCGAGTCTTCTCTCCCT 407
 QY 294 TCTCAGGAGACGCTTTGGCACTAAAGGGTCCCAACAATTTGTCTTACAAAGATACAT 353
 Db 408 CATCCGCGCGGACCGCCACTCCCGGTGCGCGCCACCGTTGTCTACAAACGATACAT 467
 QY 354 CAGCCACCGAGACATCCACATGATGCACTCAGTGGGAATCTGTGATGATTTTAC 413
 Db 468 CGCGACCGGACACCGTCCACATGAACTCTACGCGTGGCGCCACGCTCACCGAGTTGCT 527
 QY 414 TAAGTGGCTGGCAGAGAGGCTTGTGCAAAAGTGCACCTCAAGACGATAGAGTT 469
 Db 528 CAACTCTCTGGCGCGGAGAGGCTACTGCAAGGTTGAGGACACGCCCAAGGGGTGGT 583

RESULT 13

AA01559
 ID AA01559 standard; cDNA; 300 BP.
 AC AA01559;
 XX
 DT 19-MAY-2000 (first entry)

XX Human colon cancer cell line polynucleotide sequence SEQ ID NO:1550.
 DE Human; colon cancer; tumour; diagnosis; gene expression product; probe;
 KW detection; cancerous state; metastasis; identification; breast cancer;
 KW oestrogen receptor-positive breast cancer; therapy;
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.
 XX Homo sapiens.
 OS
 XX
 PN WO9558675-A2.
 XX
 PD 18-NOV-1999.
 XX
 PF 13-MAY-1999; 99WO-US010602.
 XX
 PR 14-MAY-1998; 98US-0085426P.
 PR 15-MAY-1998; 98US-0085537P.
 PR 15-MAY-1998; 98US-0085636P.
 PR 21-OCT-1998; 98US-0105234P.
 PR 27-OCT-1998; 98US-0105877P.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
 XX
 DR WPI; 2000-126369/11.
 XX
 PT Polynucleotide library used to determine cancerous states of mammalian
 PT cells.
 PT
 PS Claim 1; Page 553; 1097pp; English.

XX AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 CC libraries constructed from human colon cancer cell lines. The present
 CC invention also describes a method of detecting differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell, comprising
 CC detecting at least one differentially expressed gene product in a test

CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived. The
 CC polynucleotide sequences can be used in a method for detecting
 CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of pre-
 CC metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer
 XX
 SQ Sequence 300 BP; 114 A; 55 C; 71 G; 60 T; 0 U; 0 Other;

Query Match 17.1%; Score 171; DB 3; Length 300;
 Best Local Similarity 100.0%; Pred. No. 1.2e-36;
 Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 445 AGTGCACTGAAGACGATAGGAGTTCAGCATCAGTGAAGAAAGAAAGAAATCTTCCCAGAGC 504
 Db 130 AGTGCACTGAAGACGATAGGAGTTCAGCATCAGTGAAGAAAGAAAGAAATCTTCCCAGAGC 189
 QY 505 TCAACTCAGTCTAAAGAAAGAAAGAAAGAAATCTGCACTGATGAAATCATGGAGATT 564
 Db 190 TCAACTCAGTCTAAAGAAAGAAAGAAAGAAATCTGCACTGATGAAATCATGGAGATT 249
 QY 565 GAAGAGGAAAGAAAGAAAGAAAGTCCCGAAGACAGACTACTGGCTACAGCCTGAA 615
 Db 250 GAAGAGGAAAGAAAGAAAGAAAGTCCCGAAGACAGACTACTGGCTACAGCCTGAA 300

RESULT 14

AD084255
 ID AD084255 standard; cDNA; 1614 BP.
 AC AD084255;
 XX
 DT 21-APR-2005 (first entry)

XX Plant full length insert polynucleotide seqid 2975.
 DE
 KW plant protectant; plant growth regulant; gene therapy; plant;
 KW recombinant DNA construct; physical array; plant breeding marker;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;
 KW growth rate; cell cycle pathway; disease resistance;
 KW galactomannan production; lignin production; plant growth regulator;
 KW yield; plant growth; plant development; seed oil; protein yield;
 KW protein content; gene; ss.
 XX Unidentified.

XX US2004034888-A1.
 XX
 PN 19-FEB-2004.
 XX
 PD 28-APR-2003; 2003US-00425114.
 XX
 PF 06-MAY-1999; 99US-00304517.
 PR 05-NOV-2001; 2001US-00985678.
 XX
 XX (LIUJ/) LIU J.
 PA (ZHOU/) ZHOU Y.
 PA (KOVA/) KOVALIC D K.
 PA (SCRE/) SCREEN S E.
 PA (TABAY/) TABASKA J E.
 PA (CAOY/) CAO Y.

XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
 PI WPI; 2004-180133/17.
 DR
 XX

PT New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
XX Claim 1; SEQ ID NO 2975; 15pp; English.
XX
CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
XX Sequence 1614 BP; 399 A; 362 C; 493 G; 360 T; 0 U; 0 Other;
SQ
Query Match 17.0%; Score 170.8; DB 13; Length 1614;
Best Local Similarity 64.9%; Pred. No. 3e-36;
Matches 253; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 54 CGCCATGGGGAAGTCGGATTCTTACTCCAGGCTATCGCCACAGGATCAAGTCCAA 113
Db 76 CACGATGGGGAAGCACAGATTCTGACGCGAGGCGATCGCAACCGGATCAAGCGAA 135
QY 114 GGGGCTGCAGAGCTACGCTGGTATTCGACAGATGTCGAGAGCATGTCGGGACGAGAA 173
Db 136 GGGGCTGCAGAGCTCGGTGGTACTGTTCAGATGTCTCAGAGCATGTCGGGACGAGAA 195
QY 174 TGGCTTTAAGTGTCATTTGATGTCGCAATCTCATCAGACAACTATTGCTGCTTCAGA 233
Db 196 CGGGTTCAAGTGCCATGCTGATGTCGAGTGCACAGAGCATGCGAGTGTTCGGCAT 255
QY 234 AAATCCTCAGCAGTTTATGGAATTTTTCAGAGGAATTCGAAATGACTTTCTAGAATCT 293
Db 256 GGGCCCGGACCGGTCGTCGAGGGCTTCCTCGAGGAGTTCCTCGAGTCTCTCTCCCT 315
QY 294 TCTCAGAGACGGTTTGGCACTAAAGGGTCCACAAACATTTGCTACACGAAATACAT 353
Db 316 CATCCGCGCGGCACCGCCACTCCCGCGTCCGCGCCACCGTCTCAGACGAGTACAT 375
QY 354 CAGCCACCGAGACATCCACATGAATGCCTCAGTGGAACTCTGACTGATTTTAC 413
Db 376 CGCGGACCGGCACCATCTCAGATGAATCTCCAGCGGTGGCCACCGCTCACCAGTTCGT 435
QY 414 TAAAGTGGCTGGGACAGAAAGGCTTGTGCAA 443
Db 436 CAAAGTTCTCTGGGCGCGAGGGGTACTGTAA 465
RESULT 15
ADX47874
ID ADX47874 standard; cDNA; 1713 BP.
XX
XX ADX47874;
AC
XX
XX ADX47874;
DT 21-APR-2005 (first entry)
XX
XX Plant full length insert polynucleotide seqid 22614.
DE
XX
XX plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW

KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content; gene; ss.
XX
XX Unidentified.
OS
XX US2004034888-A1.
PN
XX 19-FEB-2004.
PD
XX 28-APR-2003; 2003US-00425114.
PF
XX 06-MAY-1999; 99US-00304517.
PR
XX 05-NOV-2001; 2001US-00985678.
PR
XX (LIU/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOY/) CAO Y.
XX
XX Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
PI
XX WPI; 2004-180133/17.
DR
XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
PT
XX
XX Claim 1; SEQ ID NO 22614; 15pp; English.
PS
XX The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
XX Sequence 1713 BP; 423 A; 422 C; 504 G; 364 T; 0 U; 0 Other;
SQ
Query Match 17.0%; Score 170.8; DB 13; Length 1713;
Best Local Similarity 64.9%; Pred. No. 3.1e-36;
Matches 253; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 54 CGCCATGGGGAAGTCGGATTCTTACTCCAGGCTATCGCCACAGGATCAAGTCCAA 113
Db 235 CACGATGGGGAAGCACAGATTCTGACGCGAAGCGATCGCAACCGGATCAAGCGAA 294
QY 114 GGGGCTGCAGAGCTACGCTGGTATTCGAGATGTCGAGAGCATGTCGGGACGAGAA 173
Db 295 GGGGCTGCAGAGCTGCGGTGGTACTGTTCAGATGTTCAGAGCATGTCGGGACGAGAA 354
QY 174 TGGCTTTAAGTGTCATTTGATGTCGGAATCTCATCAGAGACAACTATTGCTGCTTCAGA 233
Db 355 CGGGTTCAAGTGCCACTGCTGCGGATCGCAGGAGGCGAGATGTCAGGTGTTCGGCAT 414

Qy	234	AAATCCTCAGCAGTTTATGATTTATTTTTCAGAGGAATCCGAAATGACTTTCTAGAACT	293
Db	415	GGCCCCCGACCGCGTCGTCGAGGGCTTCTCCGAGGAGTTCTCGAGTCCTTCTCTCCCT	474
Qy	294	TCTCAGGAGACGCTTTTGGCACTAAAAGGGTCCACAAACATTTGTCTACAACGAATACAT	353
Db	475	CATCCGCGCGCGCACCGCCACTCCCGCGTCGCGCCACCGTCGTCTACAACGAGTACAT	534
Qy	354	CAGCCACCGAGACACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTAC	413
Db	535	CGCCGACCGGCAACAGTCCACATGAATCCACGCGTGGGCCACGCTCACCGAGTTCGT	594
Qy	414	TAAGTGGCTGGGCAGAGAGGCTTGTGCA	443
Db	595	CAAGTTCCTGGGGCGGAGGGGTACTGTAA	624

Search completed: November 26, 2005, 13:40:48
Job time : 648.701 secs

This Page Blank (uspic)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:58:53 ; Search time 5601.03 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-3
Perfect score: 1002
Sequence: 1 tgatgcagctcggtaccg.....agcatcaaatggtgttcgc 1002

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_gss1.*
10: gb_gss2.*
11: gb_gss3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	626.4	62.5	1396	4	CR595908 full-leng
2	540.8	54.0	723	3	BQ186497 UI-E-EJ1-
3	524.8	52.4	1543	4	AY609991 Sus scrof
4	510	50.9	1693	7	CN256425 170004245
5	502.2	50.1	1682	1	AV721396 AV721396
6	501	50.0	1182	10	AY412518 Homo sapi
7	501	50.0	1182	10	AY412519
8	487.4	48.6	733	5	BX104466
9	478	47.7	848	1	AL558810
10	462.2	46.1	703	1	AM035210
11	462.2	46.1	786	1	AJ819816
12	462.2	46.1	864	1	AM037867
13	455.8	45.5	871	1	AJ819723
14	455.4	45.4	543	6	CB158644 K-EST0218
15	454.6	45.4	720	7	CK301009 UI-E-EJ1-
16	451.4	45.0	754	8	DN755964 GL-Cf-127
17	447	44.6	1339	4	CR618602 full-leng
18	446.2	44.5	800	1	AM038283
19	443.4	44.3	488	3	BM689918 UI-E-EJ1-
20	441	44.0	944	1	AL522387
21	437.4	43.7	638	3	BM711163
22	436	43.5	698	1	AI650375 wa90b01.x

C 23	431.4	43.1	701	1	AI807250
C 24	426.4	42.6	820	6	CF409895
C 25	425.2	42.4	949	3	BM453332
C 26	414	41.3	581	3	BP350798
C 27	414	41.3	718	8	CX866441 HESC4_10
C 28	414	41.3	882	3	BQ221694
C 29	414	41.3	965	1	AL522388
C 30	413	41.2	552	7	CN256426
C 31	412.6	41.2	908	2	BE894105
C 32	412.4	41.2	922	6	CD389584
C 33	411	41.0	722	1	AM035948
C 34	409.4	40.9	580	3	BQ130747
C 35	408.6	40.8	679	1	AM028238
C 36	404.8	40.4	591	2	BE896845
C 37	402.2	40.1	691	1	AM028686
C 38	400.2	39.9	951	7	CO725903
C 39	400.2	39.9	988	7	CO726806
C 40	397.2	39.6	739	5	BQ782294
C 41	392.4	39.2	424	1	AV721119
C 42	391	39.0	842	3	BI862461
C 43	390	38.9	663	1	AJ660240
C 44	388.8	38.8	776	2	BF571691
C 45	388	38.7	656	7	CV029155

ALIGNMENTS

RESULT 1
CR595908
LOCUS
full-length cDNA clone CS0DJ015119 of T cells (Jurkat cell line)
Cot 10-normalized of Homo sapiens (human).
ACCESSION
CR595908.1 GI:50476715
VERSION
HTC; CNSLT_CDNA.
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1. (bases 1 to 1396)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue
Foster City, CA 94024
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)
- Web: www.genoscope.cns.fr
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
1. .1396
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DJ015119"
/tissue_type="T cells (Jurkat cell line) Cot
10-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 62.5%; Score 626.4; DB 4; Length 1396;
Best Local Similarity 76.1%; Pred. No. 4.3e-151;
Matches 942; Conservative 0; Mismatches 1; Indels 295; Gaps 2;

QY 59 TGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAGGGGC 118
Db 1 TGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAGGGGC 60
QY 119 TGCAGAAGCTACGCTGGTATGCGCAGATGTCCAGAAGCAGTCCGGGACGAGATGGCT 178
Db 61 TGCAGAAGCTACGCTGGTATGCGCAGATGTCCAGAAGCAGTCCGGGACGAGATGGCT 120
QY 179 TTAAGTGTCAATTTGATGTCCGAATCTCATCAGAGACAATTTGCTGGCTTCAGAAAATC 238
Db 121 TTAAGTGTCAATTTGATGTCCGAATCTCATCAGAGACAATTTGCTGGCTTCAGAAAATC 180
QY 239 CTCAGCAGTTTATGGATTTATTTTTCAGAGAAATTCGAAATGACTTCTAGAACTTCTCA 298
Db 181 CTCAGCAGTTTATGGATTTATTTTTCAGAGAAATTCGAAATGACTTCTAGAACTTCTCA 240
QY 299 GGAGACGCTTTGGCAGCTAAAGGCTCCACAACAACATTTGCTACACGAAATACATCAGCC 358
Db 241 GGAGACGCTTTGGCAGCTAAAGGCTCCACAACAACATTTGCTACACGAAATACATCAGCC 300
QY 359 ACCGAGACACATCCATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTTAAGT 418
Db 301 ACCGAGACACATCCATGAATGCCACTCAGTGGGAACTCTGACTGATTTTACTTAAGT 360
QY 419 GGCTGGGCAGAGAGGCTTGTGCAAA----- 444
Db 361 GGCTGGGCAGAGAGGCTTGTGCAAAAGTGGACGAGACACAAAAGGCTGGTATATTCACT 420
QY 445 ----- 444
Db 421 ACATGACAGGACCCAGAAAATATCCGCGGGCAACTGGAACTGGAGAGAAAAAGAAAGC 480
QY 445 ----- 444
Db 481 AGGACCTTGATGATGAAGAAAAAACTGCCAAATTTATTGAAGACGAAGTGAGAGAGGCC 540
QY 445 ----- 444
Db 541 TGGAAAGGAAGGAACAGGAGGTCCCTTACTTTTACGGAATTAAGCAGAGAAAAATGATGAAG 600
QY 445 ----- 444
Db 601 AGAAAGTCAGTTTAAATTTGATGAAGAGGACATGTAGCTCATCCGGAGCAAATCTTCCA 660
QY 445 -----AGTGCACTCAAGACCATAGGAAGTTCAGCATCAGTGAAC 484
Db 661 AGTCAAGTACTCTGGGACCGAGTGCATCGACATGACGATAGGAAGTTCAGCATCAGTGAAC 720
QY 485 GAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAGAAAAGAAAATCTGACAC 544
Db 721 GAAAAGAAATCTTCCAGAGCTCAACTCAGTCTAAAGAAAAGAAAAGAAAATCTGACAC 780
QY 545 TCGATCAAAATCATGGAGTTGAAGGAAAAGAAAGAACTCCCGAACAGACTACTGGC 604
Db 781 TCGATCAAAATCATGGAGTTGAAGGAAAAGAAAAGAACTCCCGAACAGACTACTGGC 840
QY 605 TACAGCCTGAAATTTATTTGAAAAATTTATAACCAAGAACTCGGAGAGAAATATCATAGA 664
Db 841 TACAGCCTGAAATTTATTTGAAAAATTTATAACCAAGAACTCGGAGAGAAATATCATAGA 900
QY 665 AAAAGGCTATTGTTAAGGAAGTAAATGACAAATATACAGCTGTGTGGAAGATGATGATT 724
Db 901 AAAAGGCTATTGTTAAGGAAGTAAATGACAAATATATACAGCTGTGTGGAAGATGATGATT 960
QY 725 CTGGAGACAGCTGAAACTTGGACCACTCATTTAGAGACAGTAAATTCAGACACAGGAA 784
Db 961 CTGGAGACAGCTGAAACTTGGACCACTCATTTAGAGACAGTAAATTCAGACACAGGAA 1020
QY 785 AAAGAATCTAGTTTAAATGGAGGCTACAGAGGAAATGAAGTACCCCTAGAAATCATCA 844
Db 1021 AAAGAATCTAGTTTAAATGGAGGCTACAGAGGAAATGAAGTACCCCTAGAAATCATCA 1080

QY 845 ATGAGAAGACTTTTTCAGTACTATCGTCAATTTGAACCTGGCCCTTTAAAGACGACGAG 904
Db 1081 ATGAGAAGACTTTTTCAGTACTATCGTCAATTTGAACCTGGCCCTTTAAAGACGACGAG 1140
QY 905 TTGAAGGAATCAATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAAAATTTGTTAAACA 964
Db 1141 TTGAAGGAATCAATATGAAGACATTTCTAAACTTGCCTGAGTTTGAAAAATTTGTTAAACA 1200
QY 965 ATACCTTTAAAATCTTAAAGCATCAAAATTTGGTTTCGC 1002
Db 1201 ATA-CATTAAAATCTTAAAGCATCAAAATTTGGTTTCGC 1237
RESULT 2
BQ186497
LOCUS
DEFINITION
723 bp mRNA linear EST 30-APR-2002
UI-E-EJ1-ajr-g-04-0-UI.r1 UI-E-EJ1 Homo sapiens cDNA clone
UI-E-EJ1-ajr-g-04-0-UI 5', mRNA sequence.
ACCESSION
BQ186497
VERSION
BQ186497.1 GI:20362048
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 723)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
PUBMED
8899548
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.reegen.com).
Seq primer: M13 REVERSE.
FEATURES
Location/Qualifiers
1..723
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EJ1-ajr-g-04-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EJ1"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EJ1 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes,
AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment,
AATCGCGCAT; optic nerve, CCATTAGGTG; retina, CCGCG;
Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This

library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN	Query Match		52.4%; Score 524.8; DB 4; Length 1543;	
	Best Local Similarity		70.7%; Pred. No. 9.9e-125;	
QY	Matches		880; Conservative 0; Mismatches 72; Indels 292; Gaps 2;	
	misc_feature		1. .1543	
DB	/organism="Sus scrofa"		/mol_type="mRNA"	
	/db_xref="taxon:9823"		/clone="Clu_5302.scr.msk.pl.Contig4"	
QY	/notes="similar to NM_012311.2 Homo sapiens KIN, antigenic determinant of recA protein homolog (mouse) (KIN)"			
	ORIGIN			
QY	47		CCTGCTGCCATGGGAAAGTCGGATTTTCTTACTCCCAAGGCTATTCGCCAAGGATCA	
	DB		25	CTGTGTCGCGATGGGGAAGTCGGATTTTCTTACCCTGGAAGGCGCATGCCCAACAGGATCA
QY	107		AGTCCAAAGGGCTGCAGAAAGCTACGCTGGTATTGGCCAGATGTGCCAGAAAGCAGTGC	166
	DB		85	AATCCAAAGGGCTTCAAGAAAGCTCGCTGGTATTGCCAGATGTGCCAGAAAGCAGTGC
QY	167		ACGAGAAATGGCTTTAAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAACTATTTG	226
	DB		145	ACGAGAAATGGCTTTAAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAACTATTTG
QY	227		CTTCCAGAAATCTCAGAGAGTTCATGGATTTATTTTCAGAGGAATTTCCGAAATGACTTC	286
	DB		205	CTTCTGAAATCTCAGAGAGTTCATGGATTTATTTTCAGAGGAATTTCCGAAATGACTTC
QY	287		TAGAATCTTCCAGAGAGCGCTTTGGCACTAAAAAGGTCCACAAACATTTGTCTACAA	346
	DB		265	TAGAATCTTCCAGAGAGCGCTTTGGCACTAAAAAGGTCCACAAACATTTGTCTACAA
QY	347		AATACATCAGCCACCGAGAGCATCCACATGAATGCCACTCAGTGGGAACTCTGACTG	406
	DB		325	AGTACATCAGCCACCGAGAGCATCCACATGAATGCCACTCAGTGGGAACTCTGACTG
QY	407		ATTTTACTAAGTGGCTGGGAGAGAGGCTTTGTGCA	443
	DB		385	ATTTCCCAAGTGGCTGGGAGAGAGGCTTTGTGCAAGTGGATGGAGACACCAAAAGGCT
QY	444		-----	443
	DB		445	GGTATATTACATAGATAGAGGCCCGGAAACAATCCCGCGCAACTCGGAACACTAGAGA
QY	444		-----	443
	DB		505	AAAAGAAGAGCAGGACTTGGACGATGAAGAAAAAACTGCCAAATTTATTCAAGAACAG
QY	444		-----	443
	DB		565	TGAGACGAGTCTGGAAGGAAGGAGCAGGAGGCCCTATTTTACGGAGTTAAGCAGAG
QY	444		-----	443
	DB		625	AAAATGAAGAAGAAAAAGTGAATTTAATTTGAACAAGGAGCATGTAGTTTCAGCAGAG
QY	444		-----	475
	DB		685	CATCTTCCAATCAAGTTCCTTTGGACCAAGTGCCTGAAGACGATGGTAGCACACGCT
QY	476		CAGTGAACGAAAGAAATCTTCCAGAGCTCAACTAGTCTAAAGAAAAAGAGAAAGA	535
	DB		745	CGGTGAACGAAAGAAATCTTCCAGAGCTCAGTCAAGTCAAAAGAAAAAGAGAAAGA
QY	536		AATCTGCATCTGGATGAATCATCTGAGATTCGAAGGAGAAAAAGAAAGAACTGCC	595
	DB		805	AGTCTGCGCTCGACGAAATCATGTGAGATTTGAAGAGGAGAGAGAGAACTGCCCG
QY	596		ACTACTGGCTACAGCCTGAAATTTATTGTGAAATTTATATACCAAGAACTGGGAGAG	655

RESULT 3	AY609991	1543 bp	mRNA	linear	HTC 31-JAN-2005
LOCUS	Sus scrofa clone Clu_5302.scr.msk.pl.Contig4, mRNA sequence.				
DEFINITION	Sus scrofa clone Clu_5302.scr.msk.pl.Contig4, mRNA sequence.				
ACCESSION	AY609991				
VERSION	AY609991.1 GI:52351561				
KEYWORDS	HTC.				
SOURCE	Sus scrofa (pig)				
ORGANISM	Sus scrofa				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae; Sus.				
AUTHORS	1 (bases 1 to 1543)				
TITLE	Jorgensen,F.G., Hobolth,A., Hornshoj,H., Bendixen,C., Fredholm,M. and Schierup/M.H.				
JOURNAL	Comparative analysis of protein coding sequences from human, mouse and the domesticated pig				
PUBMED	(er) BMC Biol. 3 (1), 2 (2005)				
AUTHORS	15679890				
TITLE	2 (bases 1 to 1543)				
JOURNAL	Hornshoj,H., Bendixen,C. and Panitz,F.				
PUBMED	Direct Submission				
AUTHORS	Submitted (28-APR-2004) Animal Breeding and Genetics, Danish Institute of Agricultural Sciences, Research Centre Foulum, Postbox				

Db 865 ATCACTGGCTCAGCCTGAAATATATAGTGAAATATATAACCAAAACTTTGGAGAGAAAT 924
Qy 656 ATCATAAGAAAAGGCTATTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAAGA 715
Db 925 ATCATAAGAAAAGGCTATTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTAAAGA 984
Qy 716 TGATTGATTTCTGGAGACAAGCTGAAACTTTGACCAGACTCATTTAGAGACAGTAATTTCCAG 775
Db 985 TGATTGACTCTGGAGACAAGCTGAAACTTTGACCAGACTCATTTAGAGACAGTAATTTCCAG 1044
Qy 776 CACCAGAAAAGAAATTTCTAGTTTAAATGGAGGCTACAGAGAAATGAAGTGACCTAG 835
Db 1045 CACCAGAAAAGAAATTTCTGCTCTGAATGGAGGCTACAGAGAAACGTGGGACCTGG 1104
Qy 836 AATCCATCAATGAGAGACATTTTTCAGCTACTATCGTCAATTTGAAACTGGCCCTTTAAAG 895
Db 1105 AATCCATCAACGAGAGACGTTTTCAGCTACTGCTAGTCAATTTGAAACTGGCCCTTTAAAG 1164
Qy 896 GACGAGAGTTGAGGAATTTCAATATGAAGACATTTTCTAACTTGCTGCTGAGTTTGAAGAT 955
Db 1165 GACGAGAGTTGAGGAATTTCAATATGAAGACATTTTCTAACTTGCTGCTGAGTTTGAAGAT 1224
Qy 956 TTGTTAAACAATACCTTTAAATCTTAAAGCATCAAAATTCGTGT 999
Db 1225 TTGCTAACAGA-CATTAATAATCTTAAGCATCAAAATTCGTGT 1267

RESULT 4
CN256425 693 bp mRNA linear EST 16-MAY-2004
LOCUS 17000424506226 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
CN256425
CN256425.1 GI:47272839
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 693)
Brandenberger R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G. J.,
Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R.,
Lebkowski, J. and Stanton, L. W.
Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
15146197
Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert Length: 693 Std Error: 0.00.
Location/Qualifiers
1. .693
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cells, embryoid bodies
derived from H1, H7 and H9 cells"
/clone_lib="GRN EB"
/note="oligo dT primed, full-length enriched cDNA library
from embryoid body outgrowths derived from hES cell lines
H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
conditions."

FEATURES
source
1. .693
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cells, embryoid bodies
derived from H1, H7 and H9 cells"
/clone_lib="GRN EB"
/note="oligo dT primed, full-length enriched cDNA library
from embryoid body outgrowths derived from hES cell lines
H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
conditions."

ORIGIN
Query Match 50.9%; Score 510; DB 7; Length 693;
Best Local Similarity 100.0%; Pred. No. 5,7e-121;
Matches 510; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 445 AGTGCACTGAAGACGATAGGAAGTTTCAGCATCAGTGAAACGAAAGAAATCTTCCAGAGC 504
Db 184 AGTGCACTGAAGACGATAGGAAGTTTCAGCATCAGTGAAACGAAAGAAATCTTCCAGAGC 243
Qy 505 TCAACTCAGTCTTAAAGAAAAGAAAAGAAAATCTGCACCTGGATGAAATCATGAGATT 564
Db 244 TCACTCAGTCTTAAAGAAAAGAAAAGAAAATCTGCACCTGGATGAAATCATGAGATT 303
Qy 565 GAAGAGAAAAGAAAAGAACTGCCCGAACAGACTACTGCTACAGCCTGAAATTTATTGTG 624
Db 304 GAAGAGAAAAGAAAAGAACTGCCCGAACAGACTACTGCTACAGCCTGAAATTTATTGTG 363
Qy 625 AAAATTATAACCAAGAAAATCTGGAGAGAAATATATAGAAAAGAGGCTATTGTTAAGAA 684
Db 364 AAAATTATAACCAAGAAAATCTGGAGAGAAATATATAGAAAAGAGGCTATTGTTAAGAA 423
Qy 685 GTAATTGACAAATATACAGCTGTTGTTGAAGATGATTGATTTCTGGAGACAGCTGAAACTT 744
Db 424 GTAATTGACAAATATACAGCTGTTGTTGAAGATGATTGATTTCTGGAGACAGCTGAAACTT 483
Qy 745 GACCACTCATTTTAGAGACAGTAATTCAGCACAGGAAAGAAATTTCTAGTTTAAAT 804
Db 484 GACCACTCATTTTAGAGACAGTAATTCAGCACAGGAAAGAAATTTCTAGTTTAAAT 543
Qy 805 GGAGCTACAGAGGAAATGAAGGTACCTAGAAATCCATCAATGAGAAAGCTTTTCAGCT 864
Db 544 GGAGCTACAGAGGAAATGAAGGTACCTAGAAATCCATCAATGAGAAAGCTTTTCAGCT 603
Qy 865 ACTATCGTCATTGAACTGGCCCTTTTAAAGAGCAGAGTTGAAGGAAATTCATATATGAA 924
Db 604 ACTATCGTCATTGAACTGGCCCTTTTAAAGAGCAGAGTTGAAGGAAATTCATATATGAA 663
Qy 925 GACATTTCTAAACTTGCCTGAGTTTGAAAA 954
Db 664 GACATTTCTAAACTTGCCTGAGTTTGAAAA 693

AV721396 682 bp mRNA linear EST 16-OCT-2000
AV721396 HTB Homo sapiens cDNA clone HTBBYC10 5', mRNA sequence.
AV721396
AV721396.1 GI:10818548
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 682)
Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L.,
Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z. and Han, Z.
Homo sapiens cDNA HTB clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203 P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
Location/Qualifiers
1. .682
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HTBBYC10"
/tissue_type="Hypothalamus"
/dev_stage="Adult"

Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
2 (bases 1 to 1182)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A., Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B., Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES
Location/Qualifiers
source
1..1182
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1182
/gene="KIN"
/locus_tag="HCM4584"
ORIGIN
Query Match 50.0%; Score 501; DB 10; Length 1182;
Best Local Similarity 100.0%; Pred. No. 1.4e-118;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 445 AGTGCACTGAAGCAGTAGAGTAGTTCAGCATCAGTGAAGAAAGAAAGAAATCTTCCAGAGC 504
Db 682 ATGTGCACTGAAGCAGTAGAGTAGTTCAGCATCAGTGAAGAAAGAAAGAAATCTTCCAGAGC 741
QY 505 TCAACTCAGCTTAAAGAAAGAAAGAAAGAAATCTGCACTGATGAAATCATGGAGATT 564
Db 742 TCACTCAGCTTAAAGAAAGAAAGAAAGAAATCTGCACTGATGAAATCATGGAGATT 801
QY 565 GAAGAGGAAAGAAAGAAATCTGCCGAAACAGACTACTGGCTACAGCCTGAAATTTATTGTG 624
Db 802 GAAGAGGAAAGAAAGAAATCTGCCGAAACAGACTACTGGCTACAGCCTGAAATTTATTGTG 861
QY 625 AAATTTATACCAAGAAATCGGAGAGAAATATCATAGAAAGAAAGCTATTGTTAAGGAA 684
Db 862 AAATTTATACCAAGAAATCGGAGAGAAATATCATAGAAAGAAAGCTATTGTTAAGGAA 921
QY 685 GTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTCTGGAGACAAGCTGAAACTT 744
Db 922 GTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTCTGGAGACAAGCTGAAACTT 981
QY 745 GACCAGACTCATTTAGAGACAGTAATTCAGCACCAAGAAAGAAATTTAGTTTAAAT 804
Db 982 GACCAGACTCATTTAGAGACAGTAATTCAGCACCAAGAAAGAAATTTAGTTTAAAT 1041
QY 805 GGAGGCTACAGAGGAATGAAGTACCCTTAGATCCATCAATGAGAGACTTTTTCAGCT 864
Db 1042 GGAGGCTACAGAGGAATGAAGTACCCTTAGATCCATCAATGAGAGACTTTTTCAGCT 1101
QY 865 ACTATCGTCATTTGAAATCGCCCTTTTAAAGGACGAGAGTTGAAGAAATTCATATGAA 924
Db 1102 ACTATCGTCATTTGAAATCGCCCTTTTAAAGGACGAGAGTTGAAGAAATTCATATGAA 1161
QY 925 GACATTTCTAAACTTGCCTGA 945
Db 1162 GACATTTCTAAACTTGCCTGA 1182
RESULT 8
BX104466/c
LOCUS
DEFINITION
BX104466 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:230344, mRNA sequence.
BX104466
ACCESSION

VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
BX104466.1 GI:27832949
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo
1 (bases 1 to 733)
Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M., Radelof,U., Schneider,D. and Korn,B.
Human Unigeneset - RZPD3
Unpublished (2003)
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGp998C17463.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Human Unigeneset - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCCACAGAAACAGCTATGAC.
FEATURES
Location/Qualifiers
source
1..733
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGp998C17463 ; IMAGE:230344"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares fetal liver spleen INFLS"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac 1; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACGGAAGAAATTAATGAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Patima Bonaldo."
ORIGIN
Query Match 48.6%; Score 487.4; DB 5; Length 733;
Best Local Similarity 99.6%; Pred. No. 4.2e-115;
Matches 499; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 502 AGCTCAACTCAGCTTAAAGAAAGAAAGAAATCTGCACTGGATGAAATCATGGAG 561
Db 733 AGCTCAACTCAGCTTAAAGAAAGAAAGAAATCTGCACTGGATGAAATCATGGAG 674
QY 562 ATTGAAGAGGAAAGAAAGAAATCTGCCGAAACAGACTACTGGCTACAGCCTGAAATTTATT 621
Db 673 ATTGAAGAGGAAAGAAAGAAATCTGCCGAAACAGACTACTGGCTACAGCCTGAAATTTATT 614
QY 622 GTGAAATTTATTAACCAAGAAACTGGGAGAGAAATATCATAGAAAGAAAGCTATTGTTAAG 681
Db 613 GTGAAATTTATTAACCAAGAAACTGGGAGAGAAATATCATAGAAAGAAAGCTATTGTTAAG 554
QY 682 GAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTCTGGAGACAAGCTGAAA 741
Db 553 GAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTCTGGAGACAAGCTGAAA 494
QY 742 CTTGACCAGACTCATTTAGAGACAGTAATTCAGCACCAAGGAAAGAAATTTAGTTTAA 801
Db 493 CTTGACCAGACTCATTTAGAGACAGTAATTCAGCACCAAGGAAAGAAATTTAGTTTAA 434

```

QY      802 AATGGAGGCTACAGAGAAATGAAGTACCTAGAAATCCATCAATGAGAGAACTTTTCA 861
      |||
Db      433 AATGGAGGCTACAGAGAAATGAAGTACCTAGAAATCCATCAATGAGAGAACTTTTCA 374
      |||
QY      862 GCTACTATCGTCATTAAGAGTGGCCCTTTAAAGAGCGAGAGTGAAGGAATTCATAT 921
      |||
Db      373 GCTACTATCGTCATTAAGAGTGGCCCTTTAAAGAGCGAGAGTGAAGGAATTCATAT 314
      |||
QY      922 GAAGACATTTCTAACTTCGCTGAGTTGAAATTTGTTAAACATACCTTTAAATCTTA 981
      |||
Db      313 GAAGACATTTCTAACTTCGCTGAGTTGAAATTTGTTAAACATTA-CATTAAATCTTA 255
      |||
QY      982 AAGCATCAATTTGCTTCGC 1002
      |||
Db      254 AAGCATCAATTTGCTTCGC 234
      |||

RESULT 9
AL558810/c      848 bp      mRNA      linear      EST 02-APR-2004
LOCUS      AL558810 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
DEFINITION      Homo sapiens cDNA clone CS0DJ015Y119 3-PRIME, mRNA sequence.
ACCESSION      AL558810
VERSION      AL558810.2 GI:31280608
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
AUTHORS      Full-length cDNA libraries and normalization
TITLE      Unpublished (2001)
JOURNAL      On Feb 15, 2001 this sequence version replaced gi:12903696.
COMMENT      Genoscope - Centre National de Sequencage
      Genoscope - Centre National de Sequencage
      2 rue Gaston Cremlieux, CP 5706 - 91057 EVRY cedex - FRANCE
      Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
      1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
      end enriched, double-strand cDNA was digested with Not I and cloned
      into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
      was normalized. Library was constructed by Life Technologies, a
      division of Invitrogen. This sequence belongs to sequence cluster
      5543.r
      For more information about this cluster, see
      http://www.genoscope.cns.fr/cdna?s=CS0DJ015AE10NP1&c=5543.r.
      Location/Qualifiers
      1..848
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="CS0DJ015Y119"
      /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
      /cell_line="JURKAT"
      /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
      10-NORMALIZED"
      /note="1st strand cDNA was primed with a NotI-oligo(dT)
      primer. Five prime end enriched, double-strand cDNA was
      digested with Not I and cloned into the Not I and EcoR V
      sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES
      source
      47.7%; Score 478; DB 1; Length 848;
      Best Local Similarity 89.1%; Pred. No. 1.2e-112;
      Matches 497; Conservative 21; Mismatches 39; Indels 1; Gaps 1;

ORIGIN
Query Match      445 AGTCACCTGAAGCAGTAGAGTTTCAGCATCAGTGAACGAAAGAAATCTTCCAGAGC 504
      |||
Db      720 AGBCACKGAAGCAGDAGGAAGTTCAGCAWCAKGAAGAAAGAAAGMCCYCCCTGNGC 661
      |||
QY      505 TCAACTCAGTCTAAGAAAGAAAGAAATCTGCACCTGGATGAATCATCGAGATT 564
      |||

```

```

Db      660 TCATWTTTTTTTAAAGTATTTTGAAGAAAGAAATCTGCACCTGGATGAATCANGGAGACC 601
      |||
QY      565 GAAGAGGAAAGAAAGAACTGCGCGAACAGACTACTGGCTACAGCCTCGAAATTTATGTG 624
      |||
Db      600 GAAGAGGAAAGAAAGAAAGAAACCGCCGACAGACTCTGCGNACGCGCCGAAATTTATGTG 541
      |||
QY      625 AAAATTTATAACCAAGAAACTGGGAGAGAAATATATATAAGAAAAAGGCTATTGTTAAGGAA 684
      |||
Db      540 AAAATTTATAACCAAGAAACTGGGAGAGAAATATATATAAGAAAAAGGCTATTGTTAAGGAA 481
      |||
QY      685 GTAATGTCAAAATATACAGCTGTTGTGAAGATGATTGTTCTGGAGACAGCTGAAACCT 744
      |||
Db      480 GTAATGTCAAAATATACAGCTGTTGTGAAGATGATTGTTCTGGAGACAGCTGAAACCT 421
      |||
QY      745 GACCGACTCATTTAGAGACAGTAATTCAGACACCAAGGAAAAAGAAATCTAGTTTTAAAT 804
      |||
Db      420 GACCGACTCATTTAGAGACAGTAATTCAGACACCAAGGAAAAAGRAATCTAGTTTTAAAT 361
      |||
QY      805 GGAGGCTACAGAGGAAATGAAGTACCTAGATCCATCAATGAGAGAACTTTTTCAGCT 864
      |||
Db      360 GGAGGCTACAGAGGAAATGAAGTACCTAGATCCATCAATGAGAGAACTTTTTCAGCT 301
      |||
QY      865 ACTATCGTCATTAAGAGTGGCCCTTTAAAGAGCGAGAGTGAAGGAATTCATATGAA 924
      |||
Db      300 ACAATCGCAAGGAGACNGGCCCGNAAAGGAGCAGAGTTGAAGGAATTCATATGAA 241
      |||
QY      925 GACATTTCTAAACTTCGCTGAGTTTGAAATTTGTTAAACATACCTTTTAAATCTTAAAG 984
      |||
Db      240 GACATTTCTAAACTTCGCTGAGTTTGAAATTTGTTAAACATTA-CATTAAATCTTAAAG 182
      |||
QY      985 CATCAATTTGCTTCGC 1002
      |||
Db      181 CATCAATTTGCTTCGC 164
      |||

RESULT 10
AL558810/c      703 bp      mRNA      linear      EST 13-JUL-2005
LOCUS      AM035210 KN-252-spleen, Bos taurus Bos taurus cDNA clone
DEFINITION      C000740112 5', mRNA sequence.
ACCESSION      AM035210
VERSION      AM035210.1 GI:70790518
KEYWORDS      EST.
SOURCE      Bos taurus (cow)
ORGANISM      Bos taurus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
      Pecora; Bovidae; Bovinae; Bos.
      1 (bases 1 to 703)
      Anderson, S.I., Taylor, R., Talbot, R., Speed, D., Law, A.S.,
      Humphray, S., Hanotte, O., Mwakaya, J. and Archibald, A.L.
      Development of cDNA and EST resources for studying host responses
      to trypanosome infection in cattle
      Unpublished (2005)
      Contact: Anderson SI
      Genomics and Bioinformatics
      Roslin Institute
      Roslin, Midlothian, EH25 9PS, UNITED KINGDOM.
      Location/Qualifiers
      1..703
      /organism="Bos taurus"
      /mol_type="mRNA"
      /db_xref="taxon:9913"
      /clone="C000740112"
      /tissue_type="spleen"
      /clone_lib="KN-252-spleen, Bos taurus"
      /note="breed: N'dama"

ORIGIN
Query Match      46.1%; Score 462.2; DB 1; Length 703;
      Best Local Similarity 89.6%; Pred. No. 1.4e-108;
      Matches 508; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

```

QY 436 TTGTGCAAAAGTGCATGAGCAGATAGGAGTTTCAGCATCAGTGAACGAAAGAAATCT 495
Db 133 TTGGGACCAAGTGCCCTGAAACGATTTGGGACCGTGGCATCAGTGAAGCGGAGGAGTCC 192
QY 496 TCCGAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCACTGGATGAAATC 555
Db 193 TCCGAGAGCTCGGCTCCGTCMAAGGAGAGAGAGAGAAAGTCTGCCCTCGATGAAATC 252
QY 556 ATGGAGATTGAAGAGAGAAAAAGAAAGAACTGCCCGAACAGACTACTGCGCTACAGCTGAA 615
Db 253 ATGGAGATTGAAGAGAGAGAAAAAGAAAGAACTGCACGACAGATTAATCTGCGCTGACGCTGAA 312
QY 616 ATTATTGTGAAAATATATACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCTATT 675
Db 313 ATCATAGTGAATAATATATACCAAAAACTTTGGAGAGAGATATCATAGAAAGAGGGGATT 372
QY 676 GTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTCTGGAGACAAG 735
Db 373 GTTAAGGAAGTAAATTGACAAATATACAGCTGTCGTAAGATGATTGACTCTGGAGACAAG 432
QY 736 CTGAAACTTGCACAGACTCATTTAGAGACAGTAATTTCCAGCCAGCAGAAAAAGAAATCTA 795
Db 433 CTGAAACTTGCACAGACTCATTTAGAGACAGTCAITTTCCAGCCAGCAGAAAAAGAAATCTT 492
QY 796 GTTTTAAATGGAGGCTACAGAGGAATGAAGTACCTAGATCCATCATCAGAGACT 855
Db 493 GTTTTAAATGGAGGCTACAGAGGAATGAAGGACCCCTAGATCCATCATCAGAGACT 552
QY 856 TTTTTCAGCTACTATCTGCTCAATTTGAAACTGGCCCTTTTAAAGGCGCAGAGTTGAAAGAAAT 915
Db 553 TTTTTCGCTACTATAGTCAITGNAACTGGCCCTTTTAAAGGCGCAGAGTTGAAAGAAAT 612
QY 916 CAATATGAAGACATTTCTAACTGCTGAGTTTGAATAATTTGTTAAACAATACCTTTTAA 975
Db 613 CAATACGAAGACATTTCTAACTGCTGAGTTTGAATAATTTGTTAGC-ATCACATTA 671
QY 976 ATCTTAAAGCATCAAAATGGTGTTCG 1002
Db 672 ATCTTAAAGCATCAAAATGGTGTTCAC 698

RESULT 11
AJ819816
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 786)
McGuire,K. and Glass,E.J.
The expanding role of microarrays in the investigation of
macrophage responses to pathogens
Vet. Immunol. Immunopathol. 105 (3-4), 259-275 (2005)
Contact: McGuire K
Genomics and Genetics
Roslin Institute
Roslin, Midlothian, EH25 9PS, UNITED KINGDOM
Single pass sequencing. Bases called and trimmed with phred
v0.020425.c. Vector identified by cross match with the -minscore 20
and -minmatch 12 options. Vector:pBluescriptII(SK+) R. Site
1:EcoRV(lost) R. Site 2:NotI Seq Primer: T7 Normalised library
constructed from pooled monocytes from Bos taurus (Holstein) and
Bos indicus (Sahiwal) cattle subjected to various stimuli,
including infection with the protozoan parasite Theileria annulata.
Location/Qualifiers
1. .786
/organism="Bos sp."
/mol_type="mRNA"

/db_xref="taxon:29061"
/clone="C0006016J08"
/tissue_type="blood"
/cell_type="bovine monocyte"
/clone_lib="KN206"
/note="Vector: pBluescriptII(SK+); Site 1: EcoRV(lost);
Site 2: NotI; Normalised library constructed from pooled
monocytes from Bos taurus (Holstein) and Bos indicus
(Sahiwal) cattle subjected to various stimuli, including
infection with the protozoan parasite Theileria annulata"
ORIGIN
Query Match 46.1%; Score 462.2; DB 1; Length 786;
Best Local Similarity 89.6%; Pred. No. 1.5e-108;
Matches 508; Conservative 0; Mismatches 58; Indels 1; Gaps 1;
QY 436 TTGTGCAAAAGTGCATGAGCAGATAGGAGTTTCAGCATCAGTGAACGAAAGAAATCT 495
Db 204 TTGGGACCAAGTGCCCTGAAACGATTTGGGACCGTGGCATCAGTGAAGCGGAGGAGTCC 263
QY 496 TCCGAGAGCTCAACTCAGTCTAAAGAAAAAGAAAAAGAAATCTGCACTGGATGAAATC 555
Db 264 TCCGAGAGCTCGGCTCCGTCMAAGGAGAGAGAGAAAGTCTGCCCTCGATGAAATC 323
QY 556 ATGGAGATTGAAGAGAGAAAAAGAAAGAACTGCCCGAACAGACTACTGCGCTACAGCTGAA 615
Db 324 ATGGAGATTGAAGAGAGAGAAAAAGAAAGAACTGCCCGACAGATTACTGCGCTGAGCTGAA 383
QY 616 ATTATTGTGAAAATATATACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCTATT 675
Db 384 ATCATAGTGAATAATATATACCAAAAACTTTGGAGAGAGATATCATAGAAAGAGGGGATT 443
QY 676 GTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTCTGGAGACAAG 735
Db 444 GTAAAGGAAGTAAATTGACAAATATACAGCTGTGCTGTAAGATGATTGACTCTGGAGACAAG 503
QY 736 CTGAAACTTGCACAGACTCATTTAGAGACAGTCAITTTCCAGCCAGCAGAAAAAGAAATCTA 795
Db 504 CTGAAACTTGCACAGACTCATTTAGAGACAGTCAITTTCCAGCCAGCAGAAAAAGAAATCTCT 563
QY 796 GTTTTAAATGGAGGCTACAGAGGAATGAAGTACCCCTAGATCCATCAATCAGAGACT 855
Db 564 GTTTTAAATGGAGGCTACAGAGGAATGAAGCACCCTTAGAATCCATCAATCAGAGACT 623
QY 856 TTTTTCAGCTACTATCTGCTCAATTTGAACTGCCCTTTTAAAGGCGCAGAGTTGAAAGAAAT 915
Db 624 TTTTCTGCTACTATAGTCAITTGAATCTGCCCTTTTAAAGGCGCAGAGTTGAAAGAAAT 683
QY 916 CAATATGAAGACATTTCTAACTGCTGAGTTTGAATAATTTGTTAAACAATACCTTTTAA 975
Db 684 CAATACGAAGACATTTCTAACTGCTGAGTTTGAATAATTTGTTAGC-ATCACATTA 742
QY 976 ATCTTAAAGCATCAAAATGGTGTTCG 1002
Db 743 ATCTTAAAGCATCAAAATGGTGTTCAC 769
RESULT 12
AM037867
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 864)
Anderson,S.I., Taylor,R., Talbot,R., Speed,D., Law,A.S.,
Humphray,S., Hanotte,O., Mwakaya,J. and Archibald,A.L.
AM037867 864 bp mRNA linear EST 13-JUL-2005
AM037867 KN-252-spleen, Bos taurus Bos taurus cDNA clone
C0007409b03 5', mRNA sequence.
AM037867
AM037867.1 GI:70793175
EST.
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 864)
Anderson,S.I., Taylor,R., Talbot,R., Speed,D., Law,A.S.,
Humphray,S., Hanotte,O., Mwakaya,J. and Archibald,A.L.


```

QY 976 ATCTTAAAGCATCAAAATTCGTGTTCCG 1002
| | | | | | | | | | | | | | | | | |
Db 152 ATCTTAAAGCATCAAAATTCGTGTTCC 126
| | | | | | | | | | | | | | | | | |

RESULT 14
Cb158644
LOCUS K-EST0218017 L18POOLin1 543 bp mRNA linear EST 29-JAN-2003
DEFINITION 5', mRNA sequence.
Cb158644
VERSION Cb158644.1 GI:28143780
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 543)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 12 row: F column: 09
High quality sequence stop: 543.

FEATURES
source
1..543
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L18POOLin1-12-F09"
/cell_line="SNU-354+Cho-CK+Choi-CK+HLK-3"
/lab_host="Top10P"
/clone_lib="L18POOLin1"
/notes="Organ: Liver; Vector: pTT3-Pac; Site_1: EcoRI;
Site_2: NotI; The library was contributed by the Soares
laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."

ORIGIN
Query Match 45.4%; Score 455.4; DB 6; Length 543;
Best Local Similarity 99.6%; Pred. No. 7.8e-107;
Matches 467; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 534 GAAATCTGCATGATGAATCATGAGATTGAAGAGGAAAAGAAAGAACTGCCCGAAC 593
| | | | | | | | | | | | | | | | | |
Db 1 GAAATCTGCATGATGAATCATGAGATTGAAGAGGAAAAGAAAGAACTGCCCGAAC 60
| | | | | | | | | | | | | | | | | |

QY 594 AGACTACTGGCTACAGCCTGAAATTTATGTAATTTATACCAAGAACTGGGAGAA 653
| | | | | | | | | | | | | | | | | |
Db 61 AGACTACTGGCTACAGCCTGAAATTTATGTAATTTATACCAAGAACTGGGAGAA 120
| | | | | | | | | | | | | | | | | |

QY 654 ATATCATAGAAAAGCGTATTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAA 713
| | | | | | | | | | | | | | | | | |
Db 121 ATATCATAGAAAAGCGTATTGTTAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAA 180
| | | | | | | | | | | | | | | | | |

QY 714 GATGATTGATCTGGAGACAAGCTGAAACTTGACCACTCAATTTAGACAGAGTAATTC 773
| | | | | | | | | | | | | | | | | |
Db 181 GATGATTGATCTGGAGACAAGCTGAAACTTGACCACTCAATTTAGACAGAGTAATTC 240
| | | | | | | | | | | | | | | | | |

QY 774 AGCACCAGAAAAGATTCCTAGTTTAAATGCGGCTACAGGGAATGAGGTACCT 833
| | | | | | | | | | | | | | | | | |

```

```

Db 241 AGCACCAGAAAAGATTCCTAGTTTAAATGAGGCTACAGAGAAATGAAGGTACCT 300
| | | | | | | | | | | | | | | | | |
QY 834 AGAATCCATCAATGAGAAGACTTTTTTCAGCTACTATCTCATTTGAAACTGGCCCTTTAAA 893
| | | | | | | | | | | | | | | | | |
Db 301 AGAATCCATCAATGAGAAGACTTTTTTCAGCTACTATCTCATTTGAAACTGGCCCTTTAAA 360
| | | | | | | | | | | | | | | | | |

QY 894 AGGAGCGAGCTTGAAGGAATTCATATGAAGACATTTCTAACTTGCCTGAGTTTGAA 953
| | | | | | | | | | | | | | | | | |
Db 361 AGGAGCGAGCTTGAAGGAATTCATATGAAGACATTTCTAACTTGCCTGAGTTTGAA 420
| | | | | | | | | | | | | | | | | |

QY 954 ATTTCTTAACATACCTTTTAAATCTTAAAGCATCAAAATTCGTGTTCCG 1002
| | | | | | | | | | | | | | | | | |
Db 421 ATTTCTTAACATACATTAATAAATCTTAAAGCATCAAAATTCGTGTTCCG 468
| | | | | | | | | | | | | | | | | |

RESULT 15
Cb301009/c
LOCUS UI-E-EJ1-ajr-g-04-0-UI.s1 UI-E-EJ1 Homo sapiens cDNA clone
DEFINITION UI-E-EJ1-ajr-g-04-0-UI 3', mRNA sequence.
Cb301009
VERSION CK301009.1 GI:39890957
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 720)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Greg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/eye.html
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
Location/Qualifiers
1..720
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EJ1-ajr-g-04-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stages="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EJ1"
/notes="Organ: eye; Vector: pTT3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EJ1 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pTT3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes,

```

AGAATCAAGA: lens, CGATTAGCGA: eye anterior segment, AATGCCGAT; optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI). TAG TISSUE=human lens TAG LIB=UI-E-EJ1 TAG_SEQ=CGATTAGCGA "

ORIGIN

Query Match	45.4%	Score	454.6;	DB	7;	Length	720;		
Best Local Similarity	98.9%	Pred. No.	1.3e-106;						
Matches	468;	Conservative	0;	Mismatches	4;	Indels	1;	Gaps	1;

Qy	530	AAAAGAAATCTGACCTGGATGAATCATGAGATTGAAGAGGAAAGAAAAGAACTGCC	589
Db	715	AAAAGAAATCTGACCTGGATGAATCATGAGATTGAAGAGGAAAGAAAAGAACTGCC	656
Qy	590	GAACAGACTACTGGCTACAGCTGAAATTAATTTGAAAAATTATAACCAAGAACTGGGAG	649
Db	655	GAACAGACTACTGGCTACAGCTGAAATTAATTTGAAAAATTATAACCAAGAACTGGGAG	596
Qy	650	AGAAATATCATAGAAAAGGCTATTCTTAAGGAAGTAATTGACAAATATACAGCTGTTG	709
Db	595	AGAAATATCATAGAAAAGGCTATTCTTAAGGAAGTAATTGACAAATATACAGCTGTTG	536
Qy	710	TGAAGATGATTGATCTGGAGACAAGCTGAAAATTGACACGACTCATTTAGACACAGTAA	769
Db	535	TGAAGATGATTGATCTGGAGACAAGCTGAAAATTGACACGACTCATTTAGACACAGTAA	476
Qy	770	TTCAGCACCAGGAAAGAAATCTAGTTTTAAATGGAGGCTACAGAGGAAATGAAGGTA	829
Db	475	TTCAGCACCAGGAAAGAAATCTAGTTTTAAATGGAGGCTACAGAGGAAATGAAGGTA	416
Qy	830	CCCTAGAAATCCATCAATGAGAAGACTTTTTCAGCTACTATCGTCAATTTGAAACTGGCCCTT	889
Db	415	CCCTAGAAATCCATCAATGAGAAGACTTTTTCAGCTACTATCGTCAATTTGAAACTGGCCCTT	356
Qy	890	TAAAGGACGAGAGTTGAAGGAAATTCATATGAAGACATTTCTAAACTTGCCTGAGTTT	949
Db	355	TAAAGGACGAGAGTTGAAGGAAATTCATATGAAGACATTTCTAAACTTGCCTGAGTTT	296
Qy	950	GAATAATTGTTAAACAATACCTTTAAATCTTTAAAGCATCAAAATGGTTCGC	1002
Db	295	GAATAATTGTTAAACAATACCTTTAAATCTTTAAAGCATCAAAATGGTTCGC	244

Search completed: November 27, 2005, 00:57:50
Job time : 5606.03 secs

This Page Blank (uspjc)

; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 379
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-379

Query Match 12.3%; Score 123.6; DB 3; Length 538;
Best Local Similarity 57.5%; Pred. No. 2.2e-26;
Matches 222; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 553 ATCATGGAGATTGAAGAGGAAAGAAAGAACTGCCGGAACAGACTACTGGCTACAGCCT 612
DB 522 ATCAAGCAGGAGAAAGCAAAAGGAGCGTGCCAAACGCAAGACTACTGGCTGCACAA 463
QY 613 GAAATATTGTGAAATATTAAACCAAGAACTGGGAGAGAAATATCATAGAAAAAGGCT 672
DB 462 GGTATCGTGTCAATTTATTTCCAAATCCATGGGCGAAAAGTTCTTCAACAAAAAGCG 403
QY 673 ATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATGATCTGGAGAC 732
DB 402 GTTGTCTCGAGCTAATTGACAGATATCAGGCGGCAAAATCAAGTTCTTGAGAGCTGGGAA 343
QY 733 AAGCTGAACCTTGACAGACTCATTTAGAGACAGTAATTCAGACACAGGAAAAAGATT 792
DB 342 AAGCTAAAAAGTGGATCAAGCTCATTTGGAGACGGTAATCCCCGCTTTGGACAAGCCTGTC 283
QY 793 CTAGTTTTAAATGGAGGCTACAGAGAAATGAAGTACCTAGAAATCCATCAATGAGAAG 852
DB 282 ATGGTGGTTAATGGCGCTTATCGGGATCCGAGGCTCTGCTAAGGAAACTGGACGAGCG 223
QY 853 ACTTTTTCAGCTACTATCGTCAATTGAAACTGGCCCTTTAAAGGACGAGAGTTGAAGGA 912
DB 222 AGATATTCAAGTCAGCGTGGAATATTGACGGTCTCTCAAAGGCAGAAATTGTAGACAAC 163
QY 913 ATTCAATATGAACATTTCTAAACT 938
DB 162 GTGCAATACGAAGATATATCTAAACT 137

RESULT 3
US-09-270-767-15661/c
; Sequence 15661, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15661
; LENGTH: 538
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15661

Query Match 12.3%; Score 123.6; DB 3; Length 538;
Best Local Similarity 57.5%; Pred. No. 2.2e-26;
Matches 222; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 553 ATCATGGAGATTGAAGAGGAAAGAAAGAACTGCCGGAACAGACTACTGGCTACAGCCT 612
DB 522 ATCAAGCAGGAGAAAGCAAAAGGAGCGTGCCAAACGCAAGACTACTGGCTGCACAA 463
QY 613 GAAATATTGTGAAATATTAAACCAAGAACTGGGAGAGAAATATCATAGAAAAAGGCT 672

DB 462 GGTATCGTGTCAAAATTTATTTCCAAATCCATGGCGGAAAAGTTCTTCAACAAAAAGCG 403
QY 673 ATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATGATCTGGAGAC 732
DB 402 GTTGTCTCGAGCTAATTGACAGATATCAGGCGGCAAAATCAAGTTCTTGAGAGCTGGGAA 343
QY 733 AAGCTGAACCTTGACAGACTCATTTAGAGACAGTAATTCAGACACAGGAAAAAGATT 792
DB 342 AAGCTAAAAAGTGGATCAAGCTCATTTGGAGACGGTAATCCCCGCTTTGGACAAGCCTGTC 283
QY 793 CTAGTTTTAAATGGAGGCTACAGAGAAATGAAGTACCTAGAAATCCATCAATGAGAAG 852
DB 282 ATGGTGGTTAATGGCGCTTATCGGGATCCGAGGCTCTGCTAAGGAAACTGGACGAGCG 223
QY 853 ACTTTTTCAGCTACTATCGTCAATTGAAACTGGCCCTTTAAAGGACGAGAGTTGAAGGA 912
DB 222 AGATATTCAAGTCAGCGTGGAATATTGACGGTCTCTCAAAGGCAGAAATTGTAGACAAC 163
QY 913 ATTCAATATGAACATTTCTAAACT 938
DB 162 GTGCAATACGAAGATATATCTAAACT 137

RESULT 4
US-09-248-796A-5938
; Sequence 5938, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248.796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 5938
; LENGTH: 825
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-5938

Query Match 7.9%; Score 79; DB 3; Length 825;
Best Local Similarity 52.8%; Pred. No. 5e-13;
Matches 220; Conservative 0; Mismatches 190; Indels 7; Gaps 2;

QY 57 CATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCCCAACAGGATCAAGTCCAAGG 116
DB 105 CATGCAAAAGGCAGAGTTTGGAAACAGCAAAATATCAATCTAAAAAGTACGAGCGCAG 164
QY 117 GCTGCAGAAAGTACGCTGGTATTCAGATGTCGAGAGCAGTCCGGGAGCAAGATGG 176
DB 165 ATTACAAAAATTGAAGTTCTATTGCGCAATTTGTCTTAAACAGTGTGAGATTCAAAT 224
QY 177 CTTTAAAGTGTATTGTATGTCGGAATCTCATCAGAGACAACT-----ATTCTGGCTTC 230
DB 225 GTTTAAAAACCAATCTATCTTACCGCTGCACATAAAGAGGTTTCGGAATATACATGA 284
QY 231 AGAAAAATCCTCAGCAGTTTATGGATTTATTTTCAGAGAAATTTCCGAATATGATTTCTAG 290
DB 285 TGGTGATAGTCCAAACTAATAGAAACCTACTCTACAAAAATCCCAAGATAAGTTTATCAA 344
QY 291 ACTTCTCAGGAGACGCTTTGGCACTAAAAGGTCACAAACACATTTGTCTACACGAATA 350
DB 345 GTTGTCTCGGAATCAACCACTGGAACCTAAAATTTATCAATGCCAATAAGTTCTACCA 404
QY 351 CATCAGCACCGAGAGACATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTT 410
DB 405 TATCCGCGAGCGAGATCATATCCATATGANTTCCACCAGATGGAGAGTCTAACTTCA 464

QY 411 TACTAAGTGGCTGGCGAGAGA-AGGCTTGTGCAAAAGTGCACCTGAAGACGATAGGAA 466
DB 465 TATAAACACATTGGGTAAATAATGGGATTGTGAAAGTGCAAACTAATGATGAGAGTAA 521

RESULT 5

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; CLONE: ptzgpt-F18
; US-08-232-463-14

Query Match 7.4%; Score 74.6; DB 2; Length 7218;
Best Local Similarity 6.2%; Pred. No. 3.4e-11;
Matches 26; Conservative 237; Mismatches 156; Indels 0; Gaps 0;

QY 406 GATTTTACTAGTGGCTGGCGAGAGAAGGCTTGTGCAAAAGTGCACCTGAAGACGATAGGA 465
DB 1445 GAATTTGGTACRR 1386
QY 466 AGTTTCAGCATCAGTGAAGCAAGAAAGAAATCTTCCAGAGCTCAACTAGTCTAAAGAAAG 525
DB 1385 RRR 1326
QY 526 AAGAAAAGAAATCTGCATCGATGAAATCATGAGATTGAAGAGAAAGAAAGAACT 585
DB 1325 RRR 1266
QY 586 GCCCGACAGACTACTGGCTACAGCTGAATATTGTGAAATTTATTAACCAAGAAACTG 645

DB 1265 RRR 1206
QY 646 CGAGAGAAATATCATAGAAAGAAAGGCTATTGTTAGGAAGTAATTCACAAATATACAGCT 705
DB 1205 RRR 1146
QY 706 GTTGTGAAGATGATTCTCGAGACAAAGCTGAAATCTTGACACGACTCATTTAGACAGA 765
DB 1145 RRR 1086
QY 766 GTAATTCAGCAGCAGGAGAAAGAAAGATTCCTAGTTTAAATGGAGGCTACAGAGCAATGA 824
DB 1085 RRR 1027

RESULT 6

US-09-949-016-13872/c
; Sequence 13872, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13872
; LENGTH: 137394
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(137394)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13872

Query Match 4.5%; Score 44.6; DB 3; Length 137394;
Best Local Similarity 53.1%; Pred. No. 0.15; Indels 0; Gaps 0;
Matches 95; Conservative 0; Mismatches 84;

QY 518 AAGAAAAGAGAAAAGAAATCTGCACCTGGATGAAATCATGGAGATTGAAGAGAAAGA 577
DB 15531 AAGAAAAGAGAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 15472
QY 578 AAGAAACTGCCGAAACAGACTACTGGCTACAGCTGAAATTTTGTGAAATTTATACCA 637
DB 15471 AAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 15412
QY 638 AAGAACTGGAGAGAAATATCATAGAAAAGAGCTATTGTTAAGGAAGTATTTCACAAA 696
DB 15411 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 15353

RESULT 7

US-09-949-016-12178/c
; Sequence 12178, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755


```

US-10-071-411A-63
; Sequence 63, Application US/10071411A
; Patent No. 6797475
; GENERAL INFORMATION:
; APPLICANT: Glenn Barnes
; APPLICANT: Joanne Meyer
; TITLE OF INVENTION: Detection of Polymorphisms in the Human
; FILE REFERENCE: MRI-021
; CURRENT APPLICATION NUMBER: US/10/071,411A
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: 60/267,515
; PRIOR FILING DATE: 2001-02-08
; PRIOR APPLICATION NUMBER: 60/314,248
; PRIOR FILING DATE: 2001-08-21
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: fastseq for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 168174
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(168174)
; OTHER INFORMATION: n = A,T,C or G
US-10-071-411A-63

Query Match          4.2%; Score 42.2; DB 3; Length 168174;
Best Local Similarity 46.6%; Pred No. 0.85;
Matches 169; Conservative 0; Mismatches 193; Indels 1; Gaps 1;

Qy      498  CCAGAGCTCAACTCAGTCTAAAGAGAAAGAGAAAGAAATCTGCACCTGGATGAAATCAT 557
Db      146251 CAAGAGTGAAGTCTGTCTCAAAAAAAAAAAAAAAAAAGAGAAAGAGAAAGAGAG 146310

Qy      558  GGAGATTGAAGAGGAGAAAGAAAGAACTGCCGAGACAGACTACTGGCTTACAGCTTGAAT 517
Db      146311 GAAAGAAAGACAGAGAGAGAGAAAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGA 146370

Qy      618  TATTGTGAAATATTAAACCAAGAACTGGAGAGAAATATCATTAAGAAAGAGCTATTGT 677
Db      146371 AAAAGAAAGAAAGAAAGAAAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAG 146430

Qy      678  TAAGGAAGTAAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTTCTGGAGACAAAGCT 737
Db      146431 AAAGAAAGAGAAAGAAAGAAAGAAAGAGAAAGAGAAAGAGAAAGAGAAAGAG 146490

Qy      738  GAACTTGACGACAGACTCATTTAGAGACAGTAAATTCAGCACAGGAGAAAGAAATTTCTAGT 797
Db      146491 AAAGAAAGAAAGAAAGAAAGAAAGAGAAAGAGAAAGAGAAAGAGAGAAAGAGAA 146549

Qy      798  TTTAATCGAGCTACAGAGGAATCGAGGTACCCCTAGAGTCCATCAATGAGAGAGACTTT 857
Db      146550 GAAAGAAAGAAAGAAAGAGAGACAGAACGAAGGAAAGAGAGAAATTTTCCCTTT 146609

Qy      858  TTC 860
Db      146610 TTC 146612

RESULT 11
US-10-071-411A-2
; Sequence 2, Application US/10071411A
; Patent No. 6797475
; GENERAL INFORMATION:
; APPLICANT: Glenn Barnes
; APPLICANT: Joanne Meyer
; TITLE OF INVENTION: Detection of Polymorphisms in the Human
; FILE REFERENCE: MRI-021
; CURRENT APPLICATION NUMBER: US/10/071,411A
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: 60/267,515

```

```
Query Match          4.2%; Score 42; DB 3; Length 601;
Best Local Similarity 55.7%; Pred. No. 0.041;
Matches 78; Conservative 1; Mismatches 61; Indels 0; Gaps 0;

QY 840 CATCAATGAGAAGACTTTTTCAGCTACTATCGTCATTGAAACTGGGCCCTTTAAAGGACG 899
DB 385 CTTCACTGGAGCTGCTTCTTCAGTTACTCTATTTCACAGCAAGTTGAACCTCTAAAGGAA 326
QY 900 CAGAGTTGAAGGAATTCATATATGAAGACATTTCTTAAACTTGGCTGAGTTTGAATAATTTGT 959
DB 325 AAGAGTTTAAACATTACAGAAATTCMATTTCTTCTTAAACTTCCAAAGATTAAACATGACT 266
QY 960 TAACAATACCTTTAAATCT 979
DB 265 CATCCAAAACGATTTTACCT 246

RESULT 13
US-09-949-016-162984/c
; Sequence 162984, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162984
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-162984

Query Match          4.2%; Score 42; DB 3; Length 601;
Best Local Similarity 55.7%; Pred. No. 0.041;
Matches 78; Conservative 1; Mismatches 61; Indels 0; Gaps 0;

QY 840 CATCAATGAGAAGACTTTTTCAGCTACTATCGTCATTGAAACTGGGCCCTTTAAAGGACG 899
DB 385 CTTCACTGGAGCTGCTTCTTCAGTTACTCTATTTCACAGCAAGTTGAACCTCTAAAGGAA 326
QY 900 CAGAGTTGAAGGAATTCATATATGAAGACATTTCTTAAACTTGGCTGAGTTTGAATAATTTGT 959
DB 325 AAGAGTTTAAACATTACAGAAATTCMATTTCTTCTTAAACTTCCAAAGATTAAACATGACT 266
QY 960 TAACAATACCTTTAAATCT 979
DB 265 CATCCAAAACGATTTTACCT 246

RESULT 14
US-09-433-699-3
; Sequence 3, Application US/09433699B
; Patent No. 6165786
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Lex M. Cowsett
; TITLE OF INVENTION: ANTISENSE MODULATION OF NUCLEOLIN EXPRESSION
; FILE REFERENCE: RTS-0109
; CURRENT APPLICATION NUMBER: US/09/433,699B
; CURRENT FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 2518
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (112)..(2235)
US-09-433-699-3

Query Match          4.2%; Score 41.6; DB 3; Length 2518;
Best Local Similarity 50.5%; Pred. No. 0.12;
Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 591 AACGACTACTGGCTTACAGCCTGAAATTTATTGTGAAAATTTATACCAAGAAACTGGGAGA 650
DB 753 ACCAGCCAAAGAAAGAAAGCTGCAAAAGTTGTTCTGTGAAAGCCAAAGACGTGCTGA 812
QY 651 GAAATATCATAGAAAAAGGCTATTGTTAAGAAAGTAATTTGACAAATATACAGCTGTTGT 710
DB 813 GGATGAAGATGAAGAAAGAGGATGATGAGGACGAGGATGACGACGACGAAAGATGATGA 872
QY 711 GAAGATGATTGATCTCGAGACAGCTGAAACTTGACCACTCATTTTAGACACAGTAAT 770
DB 873 AGATGATGATGATGAAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 932
QY 771 TCCAGCACCGAAAAAGAA 790
DB 933 AGAAGCACCTGGAAAAACGAA 952

RESULT 15
US-09-949-016-4169
; Sequence 4169, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4169
; LENGTH: 2527
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4169

Query Match          4.2%; Score 41.6; DB 3; Length 2527;
Best Local Similarity 50.5%; Pred. No. 0.12;
Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 591 AACGACTACTGGCTTACAGCCTGAAATTTATTGTGAAAATTTATACCAAGAAACTGGGAGA 650
DB 753 ACCAGCCAAAGAAAGAAAGCTGCAAAAGTTGTTCTGTGAAAGCCAAAGACGTGCTGA 812
QY 651 GAAATATCATAGAAAAAGGCTATTGTTAAGAAAGTAATTTGACAAATATACAGCTGTTGT 710
DB 813 GGATGAAGATGAAGAAAGAGGATGATGAGGACGAGGATGACGACGACGAAAGATGATGA 872
QY 711 GAAGATGATTGATCTCGAGACAGCTGAAACTTGACCACTCATTTTAGACACAGTAAT 770
DB 873 AGATGATGATGATGAAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 932
QY 771 TCCAGCACCGAAAAAGAA 790
DB 933 AGAAGCACCTGGAAAAACGAA 952
```

Search completed: November 27, 2005, 01:10:41
Job time : 210.505 secs

This Page Blank (usp1c)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 ; Search time 1155.14 Seconds
(without alignments)
7173.086 Million cell updates/sec

Title: US-09-555-529-3

Perfect score: 1002

Sequence: 1 tgaattcagctggtaccg.....agcatcaaatggtgttcgc 1002

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.Main.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	371.2	37.0	461	7	US-10-242-535A-29177
2	371.2	37.0	461	7	US-10-085-783A-29177
3	267.2	26.7	591	5	US-10-106-698-1187
4	177.6	17.7	538	7	US-10-437-963-89433
5	176	17.6	1713	8	US-10-739-930-2955
6	174.4	17.4	2046	8	US-10-425-115-157125
7	171	17.1	300	9	US-10-779-543-7646
8	170.8	17.0	1614	7	US-10-425-114-2975
9	170.8	17.0	1614	7	US-10-425-114-22614
10	170	17.0	777	7	US-10-424-599-25484
11	163.4	16.3	1241	10	US-11-097-143-41039
12	163.4	16.3	3313	10	US-11-097-143-41038
13	163.4	16.3	4394	10	US-11-097-143-41017
14	163.4	16.3	4582	10	US-11-097-143-24532
15	156.8	15.6	584	7	US-10-767-701-25587
16	126.4	12.6	549	3	US-09-991-936-1424
17	126.4	12.6	549	9	US-10-978-245-1424
18	106.8	10.7	739	7	US-10-424-599-53324
19	102.4	10.2	516	4	US-09-925-065A-549989
20	100.2	10.0	270	3	US-09-294-093B-3089
21	72.6	7.2	1608	4	US-09-925-065A-707309
22	71.2	7.1	616	7	US-10-424-599-53326
23	67	6.7	588	4	US-09-925-065A-20834

C

ALIGNMENTS

RESULT 1

US-10-242-535A-29177
; Sequence 29177, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242.535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR FILING DATE: 2002-02-28
; PRIOR FILING DATE: 2002-02-28
; PRIOR FILING DATE: 2001-07-13
; PRIOR FILING DATE: 2001-03-12
; PRIOR FILING DATE: 2001-03-12
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29177
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-29177

Query Match 37.0%; Score 371.2; DB 7; Length 461;

Best Local Similarity 91.2%; Pred. No. 1.4e+88;
Matches 405; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 51 GGTCGCCATGGGGAAGTCGGAATTTCTTACTCCCAAGGCTATCCCAACAGATCAAGTC 110
Db 1 GGTCGCCATGGGGAAGTCGGAATTTCTTACTCCCAAGGCTATCCCAACAGATCAAGTC 59
QY 111 CAAGGGGCTGCAGAAAGTACGCTGGTATGCGCATGTCGAGAGAGAGTCCCGGAGCA 170
Db 60 CAAGGGGCTGCAGAAAGTACGCTGGTATGCGCATGTCGAGAGAGAGTCCCGGAGCA 119
QY 171 GAATGGCTTTAAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAATATTGCTGGCTTC 230
Db 120 GAATGGCTTTAAAGTGTCAATGTATGTCCGAATCTCATCAGAGACAATATTGCTGGCTTC 179
QY 231 AGAAATCTCTCAGCAGTTTATGGATTATTTTTCAGAGAAATTCGGAATGACTTTCTAGA 290
Db 180 AGAAATCTCTCAGCAGTTTATGGATTATTTTTCAGAGAAATTCGGAATGACTTTCTAGA 239

QY 291 ACTTCTCAGGAGCGCTTTGGCACTAAAAGGCTCCACAACATTTGTCTTACAACGAATA 350
DB 240 ACTTCTCAGGAGCGCTTTGGCACTAAAAGGCTCCACAACATTTGTCTTACAACGAATA 299
QY 351 CATCAGCCACCGAGAGCAGATCAGATGAATGCGCTCAGTGGGAACCTCTGACTGATTT 410
DB 300 CATCAGCCACCGAGAGCAGATCAGATGAATGCGCTCAGTGGGAACCTCTGACTGATTT 359
QY 411 TACTAGTGGCTGGGCGAGAGAGCGCTTGTGCAAAAGTGCACCTGAAGACGATAGGAAGTTC 470
DB 360 TACTAGTGGCTGGGCGAGAGAGCGCTTGTGCAAAAGTGCACCTGAAGACGATAGGAAGTTC 419
QY 471 AGCATCAGTGAACGAAAGAAATC 494
DB 420 ATATTCACTATAGACAGGGACC 443

RESULT 2

US-10-085-783A-29177
; Sequence 29177, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29177
; LENGTH: 461
; TYPE: DNA
; ORGANISM: Human
US-10-085-783A-29177

Query Match 37.0%; Score 371.2; DB 7; Length 461;
Best Local Similarity 91.2%; Pred. No. 1.4e-88;
Matches 405; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 51 GCTCGCATGGGAAGTCGGATTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTC 110
DB 1 GGTGCGCATGGGAAGTCGGATTTCTTACTCCCAA-GCTATGCGCCAAACAGGATCAAGTC 59
QY 111 CAAGGGGCTGCAGAGCTACGCTGGTATTCGAGATGTGCGAGACAGTGC CGGGACGA 170
DB 60 CAAGGGGCTGCAGAGCTACGCTGGTATTCGAGATGTGCGAGACAGTGC CGGGACGA 119
QY 171 GAATGCGCTTAAAGTGTCATTTGATGTCGGAATCTCATCAGAGCAACTATTGCTGGCTTC 230
DB 120 GAATGCGCTTAAAGTGTCATTTGATGTCGGAATCTCATCAGAGCAACTATTGCTGGCTTC 179
QY 231 AGAAATCTCAGCAGTATTTATGATATTTTTCAGAGGAATTCGGAATGACTTTCTAGA 290
DB 180 AGAAATCTCAGCAGTATTTATGATATTTTTCAGAGGAATTCGGAATGACTTTCTAGA 239
QY 291 ACTTCTCAGGAGCGCTTTGGCACTAAAAGGCTCCACAACATTTGTCTTACAACGAATA 350
DB 240 ACTTCTCAGGAGCGCTTTGGCACTAAAAGGCTCCACAACATTTGTCTTACAACGAATA 299
QY 351 CATCAGCCACCGAGAGCAGATCAGATGAATGCGCTCAGTGGGAACCTCTGACTGATTT 410
DB 300 CATCAGCCACCGAGAGCAGATCAGATGAATGCGCTCAGTGGGAACCTCTGACTGATTT 359
QY 411 TACTAGTGGCTGGGCGAGAGAGCGCTTGTGCAAAAGTGCACCTGAAGACGATAGGAAGTTC 470

DB 360 TACTAGTGGCTGGGCGAGAGAGCGCTTGTGCAAAAGTGCACGAGACCAAAAGGCTGTT 419
QY 471 AGCATCAGTGAACGAAAGAAATC 494
DB 420 ATATTCACTATAGACAGGGACC 443

RESULT 3

US-10-106-698-1187
; Sequence 1187, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1187
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (587)..(587)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1187

Query Match 26.7%; Score 267.2; DB 5; Length 591;
Best Local Similarity 95.3%; Pred. No. 1.1e-60;
Matches 286; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 703 GCTGTTGTGAAGATGATTCTCGAGACAAGCTGAAACTTGACCAAGCTCATTTAGAG 762
DB 28 GCTGCAGGAATTCGCGACGAGCTCGAGACAAGCTGAAACTTGACCAAGCTCATTTAGAG 87
QY 763 ACAGTAATTCAGCACCGAGGAAAGAAATTCAGTTTTTAAATGGAGGCTACAGAGAAAT 822
DB 88 ACAGTAATTCAGCACCGAGGAAAGAAATTCAGTTTTTAAATGGAGGCTACAGAGGAAAT 147
QY 823 GAAGGTACCTTAGAATCCATCAATGAGAGACTTTTTTTCAGCTACTATCGTCATTTGAAACT 882
DB 148 GAAGGTACCTTAGAATCCATCAATGAGAGACTTTTTTTCAGCTACTATCGTCATTTGAAACT 207
QY 883 GGCCCTTTTAAAGGACGAGAGTTCGAAGAAATTCATATGAAGACATTTCTAAACTTGCC 942
DB 208 GGCCCTTTTAAAGGACGAGAGTTCGAAGAAATTCATATGAAGACATTTCTAAACTTGCC 267
QY 943 TGAGTTTGAAATTTGTTTAAACAATACCTTTTAAATCTTTAAAGCATCAAATTCGTTGCTGC 1002
DB 268 TGAGTTTGAAATTTGTTTAAACAATA-CATTAAATCTTTAAAGCATCAAATTCGTTGCTGC 326

RESULT 4

US-10-437-963-89433
; Sequence 89433, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 89433

; LENGTH: 538

; TYPE: DNA

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1

US-10-437-963-89433

Query Match 17.7%; Score 177.6; DB 7; Length 538;

Best Local Similarity 63.7%; Pred. No. 9.7e-37;

Matches 270; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 46 GCGGTGTGCGCCATGGGAGTCCGATTTTCTTACTCCCAAGGCTATCGCCACAGGATC 105

DB 112 GTCCGGAAGAGATGGGAAGCAGAGTTCTCGACCGGAGGCGATCGGAACAGATC 171

QY 106 AAGTCCAAAGGGGCTGCAGAAAGCTACGCTGGTATTGCGAGATGCGCAGAGCAGTCCCGG 165

DB 172 AAGCGAAGGGCTGCAGAAAGCTCGGTGGTACTGCGAGATGCGCAGAGCAGTCCCGC 231

QY 166 GACGAGAATGGCTTTAAGTGTCAATTGTATGTCGGAATCTCATCAGAGACAACTATTGCTG 225

DB 232 GACGAGAATGGCTTTAAGTGTCAATTGTATGTCGGAATCTCATCAGAGACAACTATTGCTG 291

QY 226 GCTTCAGAAATCTCAGCAGTTATGATTTATTTTTCAGAGGAATTCGGAATGACTTT 285

DB 292 TTCGGCAGCGCCCGCAGTCCGAGGCTTCTCCGAGGATTTCTTCGACGCTTC 351

QY 286 CTAGAACTTCTCAGGAGACGCTTGGCACTAAAAGGCTCCACAAACATTTGTCTACAAC 345

DB 352 CTACCTTGTCTCGCCGCGCCACCGACACTCCCGCATCCCGCCACCGTCTGCTACAAC 411

QY 346 GAATACATCAGCCACCGAGAGCAGATCCATGAATGCCACTCAGTGGGAACTCTGACT 405

DB 412 GAGTTTCTCAGCCAGCCACCGTCCATGAATCTCCAGCGCTGGGCGCAGCTCACC 471

QY 406 GATTTTACTAAGTGGCTGGGAGAGGCTTGGCAAAAGTGCATGAAGACATAGGA 465

DB 472 GAGTTTCTCAGTTCTCTCGGCGCGAGGCGCATGCAAGGTTGAGGACACTCCCAAAGG 531

QY 466 AGTT 469

DB 532 TGGT 535

RESULT 5

US-10-739-930-2955

; Sequence 2955, Application US/10739930

; Publication No. US20040216190A1

; GENERAL INFORMATION:

; APPLICANT: Kovalic, David K.

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH

; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT

; FILE REFERENCE: 38-21(53377)B

; CURRENT APPLICATION NUMBER: US/10/739,930

; CURRENT FILING DATE: 2003-12-18

; NUMBER OF SEQ ID NOS: 11088

; SEQ ID NO 2955

; LENGTH: 1713

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-CLUSTER7583_2

US-10-739-930-2955

Query Match 17.6%; Score 176; DB 8; Length 1713;

Best Local Similarity 63.9%; Pred. No. 4.9e-36;

Matches 266; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 54 CGCCATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAA 113

DB 171 CCGGATGGGGAAGCAGAGTTCTCGACCGGGAAGGGATCGGAACCGGATCAAGGCGAA 230

QY 114 GGGGCTGCAGAAAGCTACGCTGGTATTGCGAGATGCGCAAGAGCAGTGGCGGACGAGAA 173

DB 231 GGGGCTGCAGAAAGCTGCGGTGGTACTGCCAGATGTGTGAGAAGCAGTGGCGGACGAGAA 290

QY 174 TGGCTTTAAGTGTCAATTGTATGTCGGAATCTCATCAGAGACAACTATTTCCTGCTTCAGA 233

DB 291 CCGGTTCAAGTCCCACTGCTCGAGTCCGACCGGCGAGATCGCAGTGTTCGGCAT 350

QY 234 AAATCTCTCAGCAGTTTATGATTTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACT 293

DB 351 GGGGCGCGGACCGGCTCGTAGGGCTTCTCCGAGGAATTCCTCGAGTCTTCTCTCCCT 410

QY 294 TCTCAGGAGACGCTTTTGGCACTAAAAGGCTCCACAACAACATTTGTCTCAACGGAATACAT 353

DB 411 CATCCGCGCGCGCACCGCCACTCCCGCTCGCGCCACCGTTGTCTACAACGAGTACAT 470

QY 354 CAGCCACCGAGAGCAGATCCATGAATGCCACTCAGTGGGAACTCTGACTGATTTTAC 413

DB 471 CGCCGACCGGCGCACCGTCCACATGAATCTTACGCGTGGGCCACGCTCACCAGTTCGT 530

QY 414 TAAGTGGCTGGGAGAGGCTTGTGCAAAAGTGCATGAAGACGATAGGAAGTT 469

DB 531 CAAGTCTTGGGCGCGGAAGGGTACTGCAAGGTTGAGGACACGCCCAAGGGGTGGT 586

RESULT 6

US-10-425-115-157125

; Sequence 157125, Application US/10425115

; Publication No. US20040214272A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants

; FILE REFERENCE: 38-21(53222)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 157125

; LENGTH: 2046

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: MRT4577_74879C.1

US-10-425-115-157125

Query Match 17.4%; Score 174.4; DB 8; Length 2046;

Best Local Similarity 63.7%; Pred. No. 1.4e-35;

Matches 265; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

QY 54 CGCCATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAA 113

DB 294 CCGGATGGGGAAGCAGAGTTCTCGACCGGGAAGGGATCGGAATCGGATCAAGGCGAA 353

QY 114 GGGGCTGCAGAAAGCTACGCTGGTATTGCGAGATGCGCAAGAGCAGTGGCGGACGAGAA 173

DB 231 GGGGCTGCAGAAAGCTGCGGTGGTACTGCCAGATGTGTGAGAAGCAGTGGCGGACGAGAA 413

QY 174 TGGCTTTAAGTGTCAATTGTATGTCGGAATCTCATCAGAGACAACTATTTCGTGGCTTCAGA 233

DB 414 CCGGTTCAAGTCCCACTGCTCGAGTCCGAGTCCGACCGGCGAGATGCGAGTGTTCGGCAT 473

QY 234 AAATCTCTCAGCAGTTTATGATTTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACT 293

DB 474 GGGGCGCGGCGCGCTCGTCAGGGCTTCTCCGAGGAATTCCTCGAGTCTTCTCTCCCT 533


```
; SEQ ID NO 22614
; LENGTH: 1713
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI
US-10-425-114-22614

Query Match      17.0%; Score 170.8; DB 7; Length 1713;
Best Local Similarity 64.9%; Pred. No. 1.2e-34; Indels 0; Gaps 0;
Matches 253; Conservative 0; Mismatches 137;

QY 54 CGCCATCGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCAA 113
DB 235 CACGATGGGAGACACAGATTCCTGACGCCGAGGCGATCGCAACCGGATCAAGGCGAA 294
QY 114 GGGGCTCAGAGACTACGTGGTATTGCCAGATGTCGAAAGCAGTGGCGGAGCAGAA 173
DB 295 GGGGCTCAGAGACTACGTGGTGGTACTGTGATGTCAGATGTCGAAAGCAGTGGCGGAGCAGAA 354
QY 174 TGGGTTTAAGTGTGATGTCGGAATCTCATCAGACCAACTATTGCTGGCTTCAGA 233
DB 355 CGGGTTCAAGTGCACATGTCGAGTCCGACGAGGAGATGCGAGTGTTCGGCAT 414
QY 234 AATTCCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGMAATGACTTTCTAGAACT 293
DB 415 GGGCCCCGACCGGTCTCGAGGCTTCTCGAGGAGTTCCTCGAGTCTTCTCTCCCT 474
QY 294 TCTCAGGAGAGCGTTTGGGCTAAAGGGTCCACAAACATTTGTCTACACGAATACAT 353
DB 475 CATCCGCGCGCGACCGCCACTCCCGCTCGCGCCACCGCTGCTCTACAAACGAGTACAT 534
QY 354 CAGCCACGAGAGACATCCACATGAATGCCACTCAGTGGGAATCTCTGATGATTTTAC 413
DB 535 CGCCGACCGGACCAACGTCACATGAATCCACGCGTGGGCCACGCTCACCGAGTTTCGT 594
QY 414 TAAGTGGCTGGCAGAGAGCGTTGTGCAA 443
DB 595 CAAGTTCCTGGGCGCGAGGGGTACTGTAA 624

RESULT 10
US-10-424-599-25484
; Sequence 25484, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 25484
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1
US-10-424-599-25484

Query Match      17.0%; Score 170; DB 7; Length 777;
Best Local Similarity 57.5%; Pred. No. 1.3e-34; Indels 0; Gaps 0;
Matches 305; Conservative 0; Mismatches 225;

QY 53 TCGCATGGGGAAGTCGGATTTTCTTACTCCCAAGGCTATCGCCAAACAGGATCAAGTCCA 112
DB 135 TCGTATGGGGAAGTAAATGAGTTTCTCACCTAAAGCAATTCGCAATCGAATCAAGCAA 194
QY 113 AGGGGCTCCAAAGCTACGCTGGTATTTCGCCAGATGTGCCAAGCAGTGGCGGACGAGA 172
```

```
DB 195 AAGGATTGCAGAAGCTTCGGTGGTATTGCCAGATGTGTCCAGAGCAGTGCAGAGATGAGA 254
QY 173 ATGGCTTTAAGTGTCTCATTTATGTCGGAATCTCATCAGAGACAATATTGCTGGCTTCAG 232
DB 255 ATGGGTTTAAATGCCAATTCATGAGTGAAGGCGCCAGGCTCAATGCGAGATTTTGGAC 314
QY 233 AAAATCCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTTCCGAAATGACTTTTCTAGAAC 292
DB 315 AAAACCCACACCGGATAGTTGAGGGCTATTTCGAAAGATTTTGAGAGTACTTTTCTGGAGC 374
QY 293 TTCTCAGGAGACGCTTTGGCACTAAAGGGTCCACAAACATTTGCTTACCAACGAATACA 352
DB 375 ACATGAAGGCGCAGTCCACGATTCAGCCGTGTGGCAGCCACTGTAGTATTATAACGAATACA 434
QY 353 TCAGCCACCGAGACACATCCACATGAATGCCACTCAGTGGGAACTCTGACTGATTTTA 412
DB 435 TAAATGACAGACACCACTTCATATGAATCTTACTCAGTGGGTACGCTTACTGAGTTTG 494
QY 413 CTAAGTGGCTGGGCGACAGAAAGCTTGTGCAAAAGTGCACCTGMAAGACGATAGGAAGTTTCAG 472
DB 495 TTAAGTACTTGGTTCGAACTGGCAATGTAAAGTTGAGGAAACACCAAGGGATGGTTCA 554
QY 473 CATCAGTGAACGAAAGAAATCTTCCAGAGCTCAACTCAGTCTTAAAGAAAAGAAAGAAA 532
DB 555 TTACATATATAGATAGAGATTTCAGAAACCCCTTTTCAAGGAGAGAGATGAAGATAAGAGAA 614
QY 533 AGAAATCTGCACCTGGATGAATCATCGAGATTGAAGAGGAAAGAAAAGA 582
DB 615 TCAAGGCAGATATGGTAGATGAAGAAAGCGAGAAAGGAATCAGGAGACA 664
```

RESULT 11

```
US-11-097-143-41039
; Sequence 41039, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41039
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-41039
```

```
Query Match      16.3%; Score 163.4; DB 10; Length 1241;
Best Local Similarity 64.9%; Pred. No. 9.7e-33;
Matches 242; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
```


Db 3816 GTCTGCAGAGCTGCGCTGGTACTGCCAGATGTCGAGAGAGTGCCTGGATGAAACG 3875
Qy 176 GCTTTAAGTGTATGATGTCGGAATCTCATCAGAGACAATATTGCTGGCTTCAGAAA 235
Db 3876 GCTTCAAGTGCACACGATGAGCGAGTCCACACGCGCCAGTTGCTCTCTTTGCGGACA 3935
Qy 236 ATCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTC 295
Db 3936 ATCTTGGCAATTCCTGCACAGCTTCAGGAAGAGTTCTCCGACGGCTACATGGAGTTGC 3995
Qy 296 TCAGGAGACGCTTTGGCACTTAAAGGTCACCAACCAATTTGTCTACAAAGAAATACATCA 355
Db 3996 TGGCGCGCGGTTTCGGCACGAGCAACGACGACCAACAGATCTACCAAGGAGTACATTG 4055
Qy 356 GCACCGAGAGCAGCATCCACATGAATGCCACTAGTGGGAACTCTGACTGATTTACTA 415
Db 4056 CCACAAAGGAGCAGCATCCACATGAACCAACCGGATGGCTCACCTGTCGAGTACGTGA 4115
Qy 416 AGTGGCTGGGCG 428
Db 4116 AGTGGCTGGGCG 4128

RESULT 14
US-11-097-143-24532
; Sequence 24532, Application US/11097143
; Publication No. US2005020859A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: ARROPHILA GENES.
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24532
; LENGTH: 4582
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-24532

Query Match 16.3%; Score 163.4; DB 10; Length 4582;
Best Local Similarity 64.9%; Pred. No. 2e-32;
Matches 242; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
Qy 56 CCATGGGGAAGTCGAGTTTCTTACTCCCAAGGCTATCCCAACAGGATCAAGTCCAAAG 115
Db 3944 COATGGGTGCGCGGAGTAGTACGCCCAAGTACCTGCCCAACAGATGAATCGAAG 4003
Qy 116 GGCTGCAGAGCTACGCTGGTATGTCAGAGATGTCAGAGCAGTGCAGGACAGAAATG 175
Db 4004 GTCTGCAGAGCTGCGCTGGTACTGCCAGATGTCGAGAGCAGTGCAGGATGAAACG 4063
Qy 176 GCTTTAAGTGTATGATGTCGGAATCTCATCAGAGACAATATTGCTGGCTTCAGAAA 235

Db 4064 GCTTCAAGTGCACACGATGAGCGAGTCCACAGCGCGAGTTGCTCTCTTTGCGGACA 4123
Qy 236 ATCTCAGCAGTTTATGGATTATTTTTCAGAGGAATTCGGAATGACTTTCTAGAACTTC 295
Db 4124 ATCTTGGCAATTCCTGCACAGCTTCAGGAAGAGTTTCTCCGACGGCTACATGGAGTTGC 4183
Qy 296 TCAGGAGAGCCTTTGGCACTTAAAGGTCACCAACCAATTTGTCTACAAAGAAATACATCA 355
Db 4184 TGGCGCGCGGTTTCGGCACGAGCAACGACGCGCAACAGATCTACCAAGGAGTACATTG 4243
Qy 356 GCACCGAGAGCAGCATCCACATGAATGCCACTAGTGGGAACTCTGACTGATTTACTA 415
Db 4244 CCACAAAGGAGCAGCATCCACATGAACCGCACCGGATGGCTCACCTGTCGAGTACGTGA 4303
Qy 416 AGTGGCTGGGCG 428
Db 4304 AGTGGCTGGGCG 4316

RESULT 15
US-10-767-701-25587
; Sequence 25587, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 25587
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30963818
US-10-767-701-25587

Query Match 15.6%; Score 156.8; DB 7; Length 584;
Best Local Similarity 64.7%; Pred. No. 3.8e-31;
Matches 233; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
Qy 58 ATGGGGAAGTCGAGTTTCTTACTCCCAAGGCTATCCCAACAGGATCAAGTCCAAAGGG 117
Db 225 ATGGGGAAGCAGAGTTCTCTGACGCGCGAGCGATCCGAAACCGGATCAAGGCAAGGG 284
Qy 118 CTGCAGAGCTACGCTGGTATGCCAGATGTCGAGAGCAGTGCAGGAGAGAGTGCAGTGC 177
Db 285 CTGCAAAAGCTGCGGTGGTACTGTCAGATGTCGAGAGCAGTGCAGGAGAGTGCAGTGC 344
Qy 178 TTTAAGTGTATTGATGTCGGAATCTCATCAGAGACAATATTGCTGGCTTCAGAAAT 237
Db 345 TTCAAGTGCACCTGCATGTCGAGTCCACAGCGCGAGATGCGAGTGTTCGCGATGCG 404
Qy 238 CCTCAGCAGTTTATGGAATTTTTCAGAGAAATTCGAAATGACTTTCTAGAACTTCTC 297
Db 405 CCGACCGCGTGTGCGAGGGCTTCTCGAGAGTTCCTCGAGTCTCTCTCTCCCTCATC 464
Qy 298 AGGAGAGCTTTGGCACTTAAAGGTCACAAACATTTGCTACAAAGAAATACATCAGC 357
Db 465 CGCCGCGCGCACCGCCACTCCCGGTGCGCCCGACCGTCTGCTACACAGTACATCGCG 524
Qy 358 CACCGAGAGCAGATCCACATGAATGCCACTCAGTGGGAAACTCTGACTGATTTTACTAAG 417
Db 525 GACCGTACACAGTCCACATGAATCCACGCGCTGGGCGCACGCTCACCAGTTCGTCAAG 584

Result No.	Score	Query Match	Length	DB	ID	Description
1	44	4	437	1	US-10-821-234-402	Sequence 402, App
2	40.4	4.0	1575	7	US-11-074-176-145	Sequence 145, App
3	39.4	3.9	3254	1	US-10-793-626-4202	Sequence 4202, Ap
4	35.4	3.5	2631	7	US-11-074-176-321	Sequence 321, App
5	35.4	3.5	2649	7	US-11-074-176-87	Sequence 87, Appl
6	35	3.5	3043	1	US-10-793-626-3726	Sequence 3726, Ap
7	34	3.4	3218	1	US-10-793-626-4229	Sequence 4229, Ap
8	33.4	3.3	3939	1	US-10-793-626-3103	Sequence 3103, Ap
9	33.4	3.3	3148	1	US-10-793-626-3369	Sequence 3369, Ap
10	32.4	3.2	1167	7	US-11-074-176-323	Sequence 323, App
11	32.4	3.2	1218	7	US-11-074-176-91	Sequence 91, Appl
12	32.4	3.2	3043	1	US-10-793-626-4176	Sequence 4176, Ap
13	31.8	3.2	2078	1	US-10-793-626-4417	Sequence 4417, Ap
14	31.8	3.2	125594	1	US-10-658-986-5	Sequence 5, Appli
15	31.4	3.1	2361	1	US-10-467-962B-102	Sequence 102, App
16	31.4	3.1	340000	7	US-11-102-978-3	Sequence 3, Appli
17	31	3.1	3073	1	US-10-793-626-3581	Sequence 3581, Ap
18	31	3.1	3494	1	US-10-793-626-3650	Sequence 3650, Ap
19	31	3.1	3926	1	US-10-793-626-4300	Sequence 4300, Ap
20	31	3.1	4138	1	US-10-793-626-3995	Sequence 3995, Ap
21	30.8	3.1	2975	1	US-10-793-626-3345	Sequence 3345, Ap
22	30.8	3.1	4114	1	US-10-793-626-4184	Sequence 4184, Ap
23	30.8	3.1	43948	1	US-10-949-720-393	Sequence 393, App

RESULT 2

```
US-11-074-176-145
; Sequence 145, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Perill, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 145
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Lactobacillus acidophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1575)
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: ORF 1524; Lisk - Putative histidine kinase/may be
; OTHER INFORMATION: involved in stress response
US-11-074-176-145
```

```
Query Match          4.0%; Score 40.4; DB 7; Length 1575;
Best Local Similarity 53.9%; Pred. No. 0.071;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 613 GAAATATTGTGAAATTATAACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCT 672
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1003 GAAATGCTTGATTAACTAGAGCTGAACAAATTGATGTTCAATATCCTTATGAAGTAACA 1062

QY 673 ATTGTTAAGGAGTAATTGACAAATATACAGCTGTTGTGAAGATGTTGATTTCTGGAGAC 732
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1063 AATGTTAAGCAACAGCTAAACAGAGTTGTTTCTGATTGGCAATGGTTTCATTCAGACTTT 1122

QY 733 AAGCTGAACTTGACCAGACTCATTTAGAGACAG 766
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1123 AAGATTCAACTTGATGAAGATGATTTACCACCAG 1156
```

RESULT 3

```
US-10-793-626-4202
; Sequence 4202, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4202
; LENGTH: 3254
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
```

US-10-793-626-4202

```
Query Match          3.9%; Score 39.4; DB 1; Length 3254;
Best Local Similarity 47.7%; Pred. No. 0.2;
Matches 115; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 612 TGAATATTGTGAAATTATAACCAAGAACTGGGAGAGAAATATCATAGAAAAGGC 671
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 34 TGGAAATAAGGTCGATGTTTCATAGCATCGTTCCAGTAGGACAAGATCCACATGAATATGA 93

QY 672 TATTGTTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTGATTTCTGGAGA 731
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 94 GGTAAACCTTAAGATATTAACAGCATTAACAGATGCTGACGTTGTTATTTATAATGGTTT 153

QY 732 CAAAGCTGAACTTGACCAGACTCATTTAGAGACAGTAATTCAGCACCCAGGAAAAAGAAAT 791
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154 AAACCTAGAACTGGAAATGGTTGGTTTGAAGAAAGCACTTGACCAAGCAGGAAAAATCAAC 213

QY 792 TCTAGTTTAAATGGAGGCTACAGAGGAAGTAAGTAGTACCCTAGATCCATCAATGAGAA 851
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 214 AAAGATAAAATGTGATAGCAGCATCAAAATAAATGTTAAACCAATATATCTTAAATGGTGA 273

QY 852 G 852
Db 274 G 274
```

RESULT 4

```
US-11-074-176-321
; Sequence 321, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Perill, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 321
; LENGTH: 2631
; TYPE: DNA
; ORGANISM: Lactobacillus acidophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2631)
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: ORF 1255; Translation initiation factor
US-11-074-176-321
```

```
Query Match          3.5%; Score 35.4; DB 7; Length 2631;
Best Local Similarity 45.1%; Pred. No. 2.3;
Matches 132; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 518 AAGAAAAGAAGAAAAGAAATCTGCACCTGGATGAATCATGGAGATTGAAGAGGAAAAAGA 577
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 413 AAGAGGTCAGCTTAAATAGAGAGCTCAAAAAGCAAGAAAGAAATATCATGAGCATTTGA 472

QY 578 AAAGAACTGCCCGAACACAGACTACTGGCTACAGCCTGAAATATTGTGAAAAATTATAACCA 637
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 473 AGCATCCTTAAAAAAGAGCAGTCAAGAAAAAGATAATAAGAAAAACAGTAAAGAAAGTAATA 532

QY 638 AGAACTGGGAGAGAAATATCATAGAAAAAGGCTATTGTTAAGGAAGTAAATTGACAAAT 697
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 533 ACAAAGAGTAGACACAAAGTGGAAAGAAAGTAATCGGACCAAAAATCTTAAACCAT 592
Qy 698 ATACAGCTGTTGTAAGATGATTCTCTGGAGACAGCTGAAACTTGACCACTCAT 757
Db 593 CACCAGCCCGTTTAAAGAAAATCAACCTGCAGATAATAGGAAAGTTTACTACTCCTA 652
Qy 758 TAGAGACAGTAATTCAGGACCAAGGAAAGAAATCTTCTAGTTTAAATGGAGGC 810
Db 653 GAGTAACATCTCTGAAGACCAACCAAGGAAAGAAACGTTGTAAGGCTCGTGGC 705

RESULT 5

US-11-074-176-87
; Sequence 87, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Perill, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 87
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Lactobacillus acidophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2649)
; FEATURE: misc feature
; NAME/KEY: (0)...(0)
; LOCATION: (0)...(0)
; OTHER INFORMATION: ORF 1255; Translation initiation factor
US-11-074-176-87

Query Match 3.5%; Score 35.4; DB 7; Length 2649;
Best Local Similarity 45.1%; Pred. No. 2.3;
Matches 132; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
Qy 518 AAGAAAGACAGAAAGAAATCTGCACCTGGATGAATCATGAGATTGAGAGGAAAGA 577
Db 431 AAGAAAGTCAGCTTAATAGAGAGCTCAAAAGCAAGAAAGAAAGATATCATGAGCATTTGA 490
Qy 578 AAAGAACTGCCGACACAGACTACTGGCTACAGCTGAAATTTATGTGAAATTTATAACCA 637
Db 491 AGCATCTTAAAGAGACGCTCAGAAAGAGATATAGAAACACAGTAAGAAAGTAATA 550
Qy 638 AGAACTGGGAGAGAAATATATCAAGAAAGAGGCTATTGTTAAGGAAGTAATTGACAAAT 697
Db 551 ACAAAGAGTAGACCAACAGTGGGAAAGAAAGTAATCGGACCAAAAATCTTAAACCAT 610
Qy 698 ATACAGCTGTTGTAAGATGATTGTTCTGAGACAGCTGAAACTTGACCAAGCTCAT 757
Db 611 CACCAGCCGTTTAAAGAAATCAACCTGCAGATAATAGGAAAGAAAGTTTACTACTCCTA 670
Qy 758 TAGAGACAGTAATTCAGGACCAAGGAAAGAAATCTTCTAGTTTAAATGGAGGC 810
Db 671 GAGTAACATCTCTGAAGACCAACCAAGGAAAGAAACGTTGTAAGGCTCGTGGC 723

RESULT 6

US-10-793-626-3726/c
; Sequence 3726, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3726
; LENGTH: 3043
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3726

Query Match 3.5%; Score 35; DB 1; Length 3043;
Best Local Similarity 43.7%; Pred. No. 3.2;
Matches 203; Conservative 0; Mismatches 260; Indels 2; Gaps 1;
Qy 517 AAGAAAGAGAGAAAGAAATCTGCACCTGGATGAATCATGAGATTGAGAGGAAAG 576
Db 1332 AATTAAGAGAAATTAGAGAAATGCTAATAAGAAATAAATAATTTGTAGTAGAGAGTTGATAG 1273
Qy 577 AAGAAAGCTGCCGACAGACTACTGGCTACAGCTGAAATTTATGTGAAATTTATAACC 636
Db 1272 ATATTTTGAACATGAATCTGAGATACAACTAGTTGAAGTTTAATGAATTTGTTTTT 1213
Qy 637 AAGAAAGCTGGGAGAGAAATATATAAGAAAGAGGCTATTGTTAAGGAAGTAATTGACAAA 696
Db 1212 GTTGATAGTATAGACAGTGTCTTACGACAGAGAGGTTTCATAGAACTTTTAAAGAAATTC 1153
Qy 697 TATACAGCTGTTGTGAAGATGATTCTGAGACAGCAAGCTGAAACTTTGACCAAGCTCAT 756
Db 1152 ACAAGTTATTTTCTGATTTCACTAATTTATGTTTAAAGCGGATTTATCATCTTTGAA 1093
Qy 757 TTAGAGACAGTAATTCAGACCCAGGAGAAAGAAATTTCTAGTTTAAATGGAGGCTACAGA 816
Db 1092 ATAGAAGATTTAGTTTTCAGAGCTGAAATAAGCAATTAAGATTAAGTAAAGTAAAGTAA 1033
Qy 817 GGAATGAAGGTACCCCTAGATAATCCATCAATGAGAAGACTTTTTCAGCTACTATCGTCATT 876
Db 1032 TTAATAGATTAATTAATAAAGTTTATTAAGTAAAGTCTTATGCTATTATTAAGAACAA 973
Qy 877 GAAACTGGCCCTTTAAAGGACGACAGATTGAAAGGAATTCATATATGAAGACATTTCTAAA 936
Db 972 AGAATTGCTGGCTTAACAATTAAGAAATGATTCAAGATGGCAAGAAAGAACTTTCAAT 913
Qy 937 CTGCGCTGAGTTTGAAG--ATTGTTTAAACAATACCTTTAAATCT 979
Db 912 TTTAATAATGTTTATAATCAAGAGATTACAAGACCATTTAAACCT 868

RESULT 7

US-10-793-626-4229
; Sequence 4229, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4229
; LENGTH: 3218
; TYPE: DNA
; ORGANISM: Artificial Sequence

; FEATURE: ; OTHER INFORMATION: Description of Artificial Sequence: synthetic ; OTHER INFORMATION: nucleic acid sequence US-10-793-626-4229									
Query Match 3.4%; Score 34; DB 1; Length 3218; Best Local Similarity 48.5%; Pred. No. 6.3; Mismatches 100; Indels 0; Gaps 0; Matches 94; Conservative 0									
QY	612	TGAATATTATGTGAAATATATACCAAGAACTGGGAGAGAAATATCATAGAAAAAGGC	671						
Db	3016	TGGAATAAGTCGATGTTTCATAGCATGTTCCAGTAGCAAGATCCACATGATATCA	3075						
QY	672	TATTTGTAAGGAAGTAATTGACAAATATACAGCTGTTGTGAAGATGATTTCTGGAGA	731						
Db	3076	GGTTAAACCTAAAGATATTAAAGCATTAACAGATGCTGACGTTGTATTATATATGGTTT	3135						
QY	732	CAAGCTGAACTTGACCACTCATTTAGAGACAGTAATCCAGCACAGGAAAAAGAAAT	791						
Db	3136	AAACCTAGAACTGGAAATGGTTGGTTTGA AAAAGCACTTGACCAAGCAGGAAAAATCAAC	3195						
QY	792	TCTAGTTTTAAATG	805						
Db	3196	AAAAGATAAAATG	3209						
RESULT 8 US-10-793-626-3103 ; Sequence 3103, Application US/10793626 ; Publication No. US20050255478A1 ; GENERAL INFORMATION: ; APPLICANT: KIMMERLY, WILLIAM JOHN ; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS ; FILE REFERENCE: PU3480US ; CURRENT APPLICATION NUMBER: US/10/793,626 ; CURRENT FILING DATE: 2004-03-04 ; PRIOR APPLICATION NUMBER: 60/164,258 ; PRIOR FILING DATE: 1999-11-09 ; NUMBER OF SEQ ID NOS: 4472 ; SOFTWARE: PatentIn Ver. 2.1 ; SEQ ID NO 3103 ; LENGTH: 939 ; TYPE: DNA ; ORGANISM: Artificial Sequence ; FEATURE: ; OTHER INFORMATION: Description of Artificial Sequence: synthetic ; OTHER INFORMATION: nucleic acid sequence US-10-793-626-3103									
Query Match 3.3%; Score 33.4; DB 1; Length 939; Best Local Similarity 54.5%; Pred. No. 4.6; Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;									
QY	743	TTGACCACTCATTTAGAGACAGTAATCCAGCACAGGAAAAAGAAATCTAGTTTAA	802						
Db	71	TTGAGCATCAACATCCGGATACATTACTTCAAGAAGCTGAAATAAAAAATTCGTCTGG	130						
QY	803	ATGGAGGCTACAGAGAAATGAGGTACCTAGAAATCCATCAATGAGAAAGACTTTTTCAG	862						
Db	131	AATCAATTTAGTGTATATGATCCGAATCAGATTTGATGATGTGAATCAGCATGCAG	190						
QY	863	CTA 865							
Db	191	GTA 193							
RESULT 9 US-10-793-626-3369 ; Sequence 3369, Application US/10793626 ; Publication No. US20050255478A1 ; GENERAL INFORMATION: ; APPLICANT: KIMMERLY, WILLIAM JOHN ; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS									
Query Match 3.2%; Score 32.4; DB 7; Length 1167; Best Local Similarity 49.4%; Pred. No. 9.9; Matches 84; Conservative 0; Mismatches 86; Indels 0; Gaps 0;									
QY	513	GTCTAAAGAAAAGAGAAAAAGAAATCTGCACCTGGATGAAATCATGAGATTTGAAGAGA	572						
Db	3	GTCTAAAGAAATGCTAGAAAGCGTTCCTACATTTGAAAAAGAACAAAGGCGATTTGAACAGA	62						

Qy 573 AAAGAAAGAACTGCCGGAACAGACTACTGGCTACAGCCCTGAAATATTGTAAGAAATTAT 632
Db 63 TGTTATTGTTGATGCAATTAAAGCCGCTTTGGTAGCTGCATACAGAAAATTATATCA 122
Qy 633 AACCAAGAACTGGGAGAGAAATATCATAAAGAAAAGGCTATTGTTAAGG 682
Db 123 AGCAAAAGCTTGAAGTTCATTTTGTATGAGAGAAAGGCAATTTTAAAG 172

RESULT 11

US-11-074-176-91
; Sequence 91, Application US/11074176
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: Klenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; TITLE OF INVENTION: Stress-Related Proteins and Uses Therefore
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 91
; LENGTH: 1218
; TYPE: DNA
; ORGANISM: Lactobacillus acidophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1218)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: ORF 1259; N-utilization substance protein A
US-11-074-176-91

Query Match 3.2%; Score 32.4; DB 7; Length 1218;
Best Local Similarity 49.4%; Pred. No. 10;
Matches 84; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
Qy 513 GTCTAAAGAAAGAGAAAGAAATCTGCACTGGATGAATCATGAGATTGAAGAGGA 572
Db 54 GTCTAAAGAAATGCTAGAGCGTTTCGCTACATTGCAAAAGAAAGCAAGGCGCATTTGAACAAGA 113
Qy 573 AAAGAAAGAACTGCCGGAACAGACTACTGGCTACAGCCCTGAAATATTGTAAGAAATTAT 632
Db 114 TGTTATTGTTGATGCAATTAAAGCCGCTTTGGTAGCTGCATACAGAAAATTATATCA 173
Qy 633 AACCAAGAACTGGGAGAGAAATATCATAAAGAAAAGGCTATTGTTAAGG 682
Db 174 AGCAAAAGCTTGAAGTTCATTTTGTATGAGAGAAAGGCAATTTTAAAG 223

RESULT 12

US-10-793-626-4176/c
; Sequence 4176, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4176

; LENGTH: 3043
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4176

Query Match 3.2%; Score 32.4; DB 1; Length 3043;
Best Local Similarity 51.4%; Pred. No. 17;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
Qy 836 AATCCATCAATGAGAAGACTTTTTCAGCTACTATCTCATTTGAAGTGGCCCTTTAAAG 895
Db 2865 AATCTTGTAAATGTAATGAATCTATATACATTTCAATTTTGTATATGCTTAGAAGCGT 2806
Qy 896 GACGAGAGTTGAAGGAATTCATATGAAGACATTTTCAAACTGGCTGAGTTGAAAT 955
Db 2805 TTGAAATAAATTCGTCAGTTCCATCTTCAGTCTATTTTAAATTTGTTAAATTCATCAA 2746
Qy 956 TTGTTAAACAATACCTTTTAAATCTTA 981
Db 2745 TTGTTAAATGTTTCAGAAAATGTTA 2720

RESULT 13

US-10-793-626-4417
; Sequence 4417, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4417
; LENGTH: 3078
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4417

Query Match 3.2%; Score 31.8; DB 1; Length 3078;
Best Local Similarity 44.1%; Pred. No. 25;
Matches 132; Conservative 0; Mismatches 167; Indels 0; Gaps 0;
Qy 596 ACTACTGGCTACAGCCCTGAAATATTGTTGAAAATTATTAACAAGAACTGGGAGAGAAAT 655
Db 660 ACTACTTGTATTAGATTCTTATTGTTAAATTTAACTATAAAACACAGGCTGTAAG 719
Qy 656 ATCATAAAGAAAAGGCTATTGTTAAGGAAGTAATGCAAAATATACAGTGTGTTGTAAGA 715
Db 720 ATAAAAAATCATCTGTAAACAATACATCGGCTAGTCAATCGCCATTCGATGTTTGAAG 779
Qy 716 TGATTGATTCTGGAGACAAGCTGAAACTTGACCAGACTCATTTAGAGACAGTAATCCAG 775
Db 780 CAATGGGTGTAAGAAAATATTAAACATCTAGATGCTTGTATTACTAGACTAGAGTAG 839
Qy 776 CACCAGAAAAGAAATTTCTAGTTTAAATCGAGGCTACAGAGGAATGAAGGTACCCCTAG 835
Db 840 AAGTAATGAAAATCAAAAGTTGATGTCGCGCCCTTAAATCATTTAGGTGCTTCAAGAG 899
Qy 836 AATCCATCAATGAGAGACTTTTTCAGCTACTATCGTCATTGAAACTGGCCCTTTAAA 894
Db 900 TACTTGAAGTCGGTAACAATATGCAAGCTATTTTCGGTCTCTAAATCTGATCAATCAAA 958

RESULT 14
US-10-658-986-5/c
; Sequence 5, Application US/10658986
; Publication No. US20050255476A1
; GENERAL INFORMATION:
; APPLICANT: Clark, Abbot F.
; TITLE OF INVENTION: Methods for Diagnosing Glaucoma and Discovering Anti-Glaucoma Drugs
; FILE REFERENCE: 1581 US F
; CURRENT APPLICATION NUMBER: US/10/658,986
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: PCT/US97/21054
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: USSN 60/033,227
; PRIOR FILING DATE: 1996-12-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 125594
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-658-986-5

Query Match 3.2%; Score 31.8; DB 1; Length 125594;
Best Local Similarity 46.9%; Pred. No. 1.9e+02;
Matches 99; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 497 CCAGAGCTCAACTCAGTCTAAAGAAAAGAAAGAAAGAAATCTGCACCTGGATGAAATCA 556
DB 38437 CACTGTGCCACCTAGGCTGTAAATGACATATAAATCTCGGTGGCAACCACTAAGAA 38378

QY 557 TGGAGATTGAAGAGGAAAAGAAAAGAACTGCCGAAACAGACTACTGGCTACAGCCCTGAAA 616
DB 38377 AATACTAAAACATACAAAGAAATGACAAAGAAATTAATAATCGTTACACTAGAAA 38318

QY 617 TTATTGTGAAATATTAACCAAGAACTGGGAGAGAAATATCATAGAAAAGGCTATTG 676
DB 38317 ATATGTAATACAAAGAAAGAGGTAGTAAATGGAAGATTTAAGAAACAAAGAGACACAAA 38258

QY 677 TTAAGGAAGTAATTGACAAATATACAGCTGT 707
DB 38257 TAAAGGAACAAATGCGCAAAATATCAGAAGT 38227

RESULT 15
US-10-467-962B-102
; Sequence 102, Application US/10467962B
; Publication No. US20050246784A1
; GENERAL INFORMATION:
; APPLICANT: Plesch, Gunmar
; APPLICANT: Blau, Astrid
; APPLICANT: Daeschner, Klaus
; APPLICANT: Klein, Mathieu
; TITLE OF INVENTION: Identification of Herbicidally Active Substances
; FILE REFERENCE: 2000.857
; CURRENT APPLICATION NUMBER: US/10/467,962B
; CURRENT FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: PCT/EP02/01466
; PRIOR FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 102
; LENGTH: 2361
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2361)
; OTHER INFORMATION: ORF F4F7.26, Linie 246473
US-10-467-962B-102

Query Match 3.1%; Score 31.4; DB 1; Length 2361;
Best Local Similarity 51.0%; Pred. No. 28;

Matches 74; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 601 TGGCTACAGCCCTGAAATTAATTGTGAAAAATTATAACCAAGAAACTGGGAGAGAAAATATCAT 660
DB 490 TGGGCCGATCCTAAGATTGTAAAGAACTCATAGATGAGAAAAATGTATGAGCTGCTTGGT 549

QY 661 AAGAAAAAGGCTATTGTTAAAGAGAGTAATTGACAAATATACAGCTGTTGTGAAGATGATT 720
DB 550 GAGAAAAACCGCTGCTGATAATGAAAAACCTACAAAAAGAGGAGAGAGAGAGAGCCT 609

QY 721 GATTCTGGAGACAAGCTGAAACTTG 745
DB 610 GCCAAAGTCGAGGAAAAGAAAGCTG 634

Search completed: November 27, 2005, 02:44:30
Job time : 354.647 secs

Result No.	Score	Query Match	Length	DB	ID	Description	
1	22	100.0	22	6	AX003325	Sequence	
2	22	100.0	1002	6	AX003310	Sequence	
3	22	100.0	1117	6	CQ722054	Sequence	
4	22	100.0	1528	8	AX003308	Sequence	
5	22	100.0	1528	8	HSAJ5273	Homo sapi	
C 6	22	100.0	23065	8	HUMATP5G	DJ005273	Homo sapien
C 7	22	100.0	174361	8	ALJ58044	Human DNA	
C 8	22	100.0	194563	14	ALJ391687	Homo sapien	
C 9	18.4	83.6	7111	14	ACQ17711	Drosophila	
10	18.4	83.6	166348	14	CR847851	Danio rerio	
C 11	18.4	83.6	166906	5	AC146480	Zebrafish	
C 12	18.4	83.6	170597	5	AL929558	Drosophila	
13	18.4	83.6	177953	2	AC018491	Drosophila	
14	18.4	83.6	237848	2	AE002611	Drosophila	
C 15	17.8	80.9	460	1	AY525807	Streptococcus	
C 16	17.8	80.9	889	5	BX931016	Gallus gallus	
C 17	17.8	80.9	2169	3	CR387275	Gallus gallus	
C 18	17.8	80.9	6584	3	AY236226	Uncultured	

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Angulo-Mora, J.F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 3 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)
FEATURES source
Location/Qualifiers
1..1002
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 22; DB 6; Length 1002;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAAAGTGATCGCTGCCGTGGT 22
|||||
Db 32 AGAAAGTGATCGCTGCCGTGGT 53
RESULT 3
LOCUS CQ722054 1117 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 7988 from Patent WO02068579.
ACCESSION CQ722054
VERSION CQ722054.1 GI:42282911
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 7988 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES source
Location/Qualifiers
1..1117
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 22; DB 6; Length 1117;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAAAGTGATCGCTGCCGTGGT 22
|||||
Db 67 AGAAAGTGATCGCTGCCGTGGT 88
RESULT 4
AX003308
LOCUS AX003308 1528 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 1 from Patent WO9929845.
ACCESSION AX003308
VERSION AX003308.1 GI:9927125
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.
REFERENCE 1
AUTHORS Angulo-Mora, J.F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 1 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)
FEATURES source
Location/Qualifiers
1..1528
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 22; DB 6; Length 1528;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAAAGTGATCGCTGCCGTGGT 22
|||||
Db 40 AGAAAGTGATCGCTGCCGTGGT 61
RESULT 5
LOCUS HSAJ5273 1528 bp mRNA linear PRI 13-DEC-2000
DEFINITION Homo sapiens mRNA for Kin17 protein.
ACCESSION AJ005273
VERSION AJ005273.1 GI:3850703
KEYWORDS KIN17 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Kannouche, P., Mauffrey, P., Pinon-Lataillade, G., Mattei, M.G.,
Sarasin, A., Daya-Grosjean, L. and Angulo, J.F.
TITLE Molecular cloning and characterization of the human KIN17 cDNA
encoding a component of the UVC response that is conserved among
metazoans
JOURNAL Carcinogenesis 21 (9), 1701-1710 (2000)
PUBMED 10984102
REFERENCE 2
AUTHORS Mauffrey, P.
TITLE Direct Submission
JOURNAL Submitted (08-APR-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 -
Fontenay aux Roses Cedex, 92265, FRANCE
REMARK Revised by [3]
REFERENCE 3 (bases 1 to 1528)
AUTHORS Mauffrey, P.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1998) Mauffrey P., LGR, CEA-DSV-DRR, BP 6 -
Fontenay aux Roses Cedex, 92265, FRANCE
COMMENT On Nov 7, 1998 this sequence version replaced gi:3046739.
Kannouche P., Mauffrey P., Pinon-Lataillade G., Biard D.S.F.,
Angulo J.F.
Sequences codant pour la Proteine kin17 et leurs Applications -
French Patent Nr 97 15536 1:1-50(1997).
FEATURES source
Location/Qualifiers
1..1528
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="male"
/tissue type="testis"
1..1528
/gene="Kin17"
66..1247
/gene="Kin17"
/codon_start=1
/protein_id="CAA06462.1"
/db_xref="GI:3850704"


```

/db|xref="GOA:O60870"
/db|xref="UniProt/T-EMBL:O60870"
/translation="MGKSDFTTPRAIANRIKSLGLOKLRWYQCMQKQCRDENGFKCH
CMSSHQRLALLSENQQPMDFSEFRNDFLELLRRPGTKRVHNNIVNYEYISHR
EHIHNNATQWETTLTDFTKWLGREGCKVDPTPKGWIYIQYIDRDPETIRQLELKKKK
ODLDEDEKTAPEEQVRRGLGKEQVPTFTLSRENDEKVTNLSKGCSSSGAT
SSKSTSLGPKALTKTIGSSAYKRKESQSSSTQSKKKKSAIDIMEIEEKKRTAR
TDYWLQPEIIVKIIITKLGKRYHKKKAIKVEVIDKYTAVMKIDSGDKLDDQTHLET
VIPAPGRKILVLNGYRGNEGTLSEINEKFTSATIVTETGPKRRRVEGIQYEDISK
A"
/notes="L(liver)-type ATP synthase gamma-subunit"
/codon_start=1
/product="ATP synthase gamma-subunit"
/protein_id="BAA03995.1"
/db_xref="GI:665583"
/translation="MFSRAGVAGLSAWTLQPOWIOVRNMATLKDITRLKSIKNIOKI
TKSMKVAAYAKARERELKPARIVGLGSLALYERADIKGPEDKKKHLLIGVSSDRGL
CGAHSSTAKQMSKSEVATLTAAKGVMLVIGDKIRGILYRTHSDQFLVAFKVRKRP
PTFGDASVIALLELLNSGYEFDEGSIIIFNKRFSVISYKTEKPIFSLNTVASADMSIY
DDIDADVLQNYQVNLANIYYSLKSTTSQSRMTAMDNASKNASEMIDKLTFTN
RTRQAVITKELIIEISGAALD"
4275..11168
/number=1
11169..11203
/number=2
11204..12113
/number=2
12114..12245
/number=3
12246..14056
/number=3
14057..14261
/number=4
14262..14838
/number=4
14839..14982
/number=5
14983..15093
/number=5
/cons_splices=(5'site:no, 3'site:yes)
15094..15158
/number=6
15159..17328
/number=6
17329..17484
/number=7
17485..17816
/number=7
17817..17913
/number=8
17914..22027
/number=8
22028..22064
/number=9
22065..22712
/number=9
22713..22869
/number=10
22835..22840
22849..22854
polyA_signal
polyA_signal
ORIGIN
Query Match 100.0%; Score 22; DB 8; Length 1528;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AGAAAGTCATCGCTGCCGTGGT 22
|||||
Db 40 AGAAAGTCATCGCTGCCGTGGT 61
|||||

RESULT 6
HUMATPSG 23065 bp DNA linear PRI 13-FEB-2003
DEFINITION Homo sapiens gene for ATP synthase gamma-subunit, complete cds.
ACCESSION D16561
VERSION D16561.1 GI:468446
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 23065)
Matsuda,C., Endo,H., Ohta,S. and Kagawa,Y.
Gene structure of human mitochondrial ATP synthase gamma-subunit.
Tissue specificity produced by alternative RNA splicing
J. Biol. Chem. 268 (33), 24950-24958 (1993)
PUBMED 8227057
REFERENCE 2 (bases 1 to 23065)
Kagawa,Y.
Direct Submission
Submitted (23-JUN-1993) Yasuo Kagawa, Jichi Medical School,
Department of Biochemistry, 3311-1 Yakushiji, Minamikawachi-machi,
Tochigi 329-0498, Japan (E-mail:ykagawa@ddj.nig.ac.jp,
Tel:81-285-44-2111(ex.3149), Fax:81-285-44-1827)
Location/Qualifiers
1..23065
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db|xref="taxon:9606"
/tissue_type="blood"
/clone_lib="lambda dash II"
/notes="clones hATPG[2, 5, 21 and 23]"
4188..4274
/number=1
Join(4219..4274,11169..11203,12114..12245,14057..14261,
14839..14982,15094..15158,17329..17484,17817..17913,
22713..22716)
/notes="H(heart)-type ATP synthase gamma-subunit"
/codon_start=1
/product="ATP synthase gamma-subunit"
/protein_id="BAA03994.1"
/db|xref="GI:665584"
/translation="MFSRAGVAGLSAWTLQPOWIOVRNMATLKDITRLKSIKNIOKI
TKSMKVAAYAKARERELKPARIVGLGSLALYERADIKGPEDKKKHLLIGVSSDRGL
CGAHSSTAKQMSKSEVATLTAAKGVMLVIGDKIRGILYRTHSDQFLVAFKVRKRP
PTFGDASVIALLELLNSGYEFDEGSIIIFNKRFSVISYKTEKPIFSLNTVASADMSIY
DDIDADVLQNYQVNLANIYYSLKSTTSQSRMTAMDNASKNASEMIDKLTFTN
RTRQAVITKELIIEISGAAL"
Join(4219..4274,11169..11203,12114..12245,14057..14261,
14839..14982,15094..15158,17329..17484,17817..17913,
22028..22034)
CDS
LOCUS AL158044
DEFINITION Human DNA sequence from clone RP11-264C14 on chromosome 10 Contains
the 5' end of the gene for a novel protein (MGC10848), the ITH2
gene for inter-alpha (globulin) inhibitor, H2 polypeptide (H2P),
the KIN gene for KIN, antigenic determinant of reca protein homolog
(mouse), the 3' end of the ATP5C1 gene for ATP synthase, H+
transporting, mitochondrial F1 complex, gamma polypeptide 1
(ATP5C, ATP5C1), a novel gene and two CpG islands, complete
sequence.
ACCESSION AL158044

```

VERSION	AL158044.17	GI:16214543
KEYWORDS	HTG; ATP5C; ATP5C1; ATP5C11; CpG island; H2P; ITIH2; KIN; MGCL0848.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Euarchontoglires; Primates; Catarrhini; Homidae; Homo.	
REFERENCE	1 (bases 1 to 174361)	
AUTHORS	Lovell, J.	
TITLE	Direct Submission	
JOURNAL	Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk	
COMMENT	<p>Clone requests: clonerequest@sanger.ac.uk</p> <p>On Oct 17, 2001 this sequence version replaced gi:15131961.</p> <p>The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:</p> <p>En: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at</p> <p>http://www.sanger.ac.uk/projects/C_elegans/wormpep</p> <p>This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at</p> <p>http://www.sanger.ac.uk/HGP/Chr10</p> <p>RP11-264C14 is from the library RPCT-11.1 constructed by the group of Pieter de Jong. For further details see</p> <p>http://www.chori.org/bacpac/home.htm</p> <p>VECTOR: pBACe3.6</p> <p>----- Genome Center</p> <p>Center: Wellcome Trust Sanger Institute</p> <p>Center code: SC</p> <p>Web site: http://www.sanger.ac.uk</p> <p>Contact: vegas@sanger.ac.uk</p> <p>-----</p> <p>This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.</p>	
FEATURES	<p>1..174361 Location/Qualifiers</p> <p>1..174361 /organism="Homo sapiens"</p> <p>/mol_type="genomic DNA"</p> <p>/db_xref="taxon:9606"</p> <p>/chromosome="10"</p> <p>/clone="RP11-264C14"</p> <p>/clone_lib="RPCT-11.1"</p>	
misc_feature	<p>1</p> <p>/note="Clone left end: RP11-264C14"</p> <p>join(complement(3206..3285),</p> <p>complement(AL355374.14:26760..81619))</p>	
gene	<p>/gene="ITIH5"</p>	
mRNA	<p>/locus_tag="RP11-676E7.1-003"</p> <p>join(complement(3206..3285),</p> <p>complement(AL355374.14:81450..81619),</p> <p>complement(AL355374.14:80319..80435),</p> <p>complement(AL355374.14:50238..50406),</p> <p>complement(AL355374.14:44092..44401),</p> <p>complement(AL355374.14:40790..41349),</p> <p>complement(AL355374.14:37479..37532),</p> <p>complement(AL355374.14:34005..34121),</p> <p>complement(AL355374.14:30367..30744),</p> <p>complement(AL355374.14:23606..27721))</p>	
	<p>/gene="ITIH5"</p>	
	<p>/locus_tag="RP11-676E7.1-001"</p> <p>/product="inter-alpha (globulin) inhibitor H5"</p> <p>/note="match: ESTs: BE386984.1 BE387282.1 BG333950.1 BG476698.1 BQ225649.1</p> <p>match: cDNAs: BC004282.1"</p>	
CDS	<p>join(complement(3206..3215),</p> <p>complement(AL355374.14:81450..81619),</p> <p>complement(AL355374.14:80319..80435),</p>	
	<p>complement(AL355374.14:50238..50406),</p> <p>complement(AL355374.14:44092..44401),</p> <p>complement(AL355374.14:40790..41349),</p> <p>complement(AL355374.14:37479..37532),</p> <p>complement(AL355374.14:34005..34121),</p> <p>complement(AL355374.14:30367..30744),</p> <p>complement(AL355374.14:23606..27721))</p>	
	<p>/gene="ITIH5"</p>	
	<p>/locus_tag="RP11-676E7.1-001"</p> <p>/product="inter-alpha (globulin) inhibitor H5"</p> <p>/note="match: ESTs: AAL34750.1 AA367857.1 AA367892.1 A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1 AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1 AL550297.1 AL550915.1 AL554626.1 AL5568595.1 AL5569892.1 AL557037.1 AL574092.1 AL574207.1 AL574644.1 AL575373.1 AL577265.1 AW473748.1 BE893192.1 BE896223.1 BE898209.1 BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1 BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG326634.1 BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1 BG767953.1 B1040702.1 B1553925.1 B1559648.1 BM016522.1 BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1</p>	
	<p>complement(25552..25715), complement(24379..24480),</p> <p>complement(20853..21103),</p> <p>complement(AL355374.14:81450..81619),</p> <p>complement(AL355374.14:80319..80435),</p> <p>complement(AL355374.14:50238..50406),</p> <p>complement(AL355374.14:44092..44401),</p> <p>complement(AL355374.14:40790..41349),</p> <p>complement(AL355374.14:37479..37532),</p> <p>complement(AL355374.14:34005..34121),</p> <p>complement(AL355374.14:30367..30744),</p> <p>complement(AL355374.14:23606..27721))</p>	
	<p>/gene="ITIH5"</p>	
	<p>/locus_tag="RP11-676E7.1-001"</p> <p>join(complement(50428..50596), complement(39256..39300),</p> <p>complement(25552..25715), complement(24379..24480),</p> <p>complement(20853..21103),</p> <p>complement(AL355374.14:81450..81619),</p> <p>complement(AL355374.14:80319..80435),</p> <p>complement(AL355374.14:50238..50406),</p> <p>complement(AL355374.14:44092..44401),</p> <p>complement(AL355374.14:40790..41349),</p> <p>complement(AL355374.14:37479..37532),</p> <p>complement(AL355374.14:34005..34121),</p> <p>complement(AL355374.14:30367..30744),</p> <p>complement(AL355374.14:23606..27721))</p>	
	<p>/gene="ITIH5"</p>	
	<p>join(complement(50428..50596), complement(39256..39300),</p> <p>complement(25552..25715), complement(24379..24480),</p> <p>complement(20853..21103),</p> <p>complement(AL355374.14:81450..81619),</p> <p>complement(AL355374.14:80319..80435),</p> <p>complement(AL355374.14:50238..50406),</p> <p>complement(AL355374.14:44092..44401),</p> <p>complement(AL355374.14:40790..41349),</p> <p>complement(AL355374.14:37479..37532),</p> <p>complement(AL355374.14:34005..34121),</p> <p>complement(AL355374.14:30367..30744),</p> <p>complement(AL355374.14:23606..27721))</p>	
	<p>/gene="ITIH5"</p>	
	<p>/locus_tag="RP11-676E7.1-001"</p> <p>/product="inter-alpha (globulin) inhibitor H5"</p> <p>/note="match: ESTs: AAL34750.1 AA367857.1 AA367892.1 A1089902.1 A1090076.1 A1339610.1 A1417871.1 A14079707.1 AL540791.1 AL543337.1 AL544674.1 AL544702.1 AL550125.1 AL550297.1 AL550915.1 AL554626.1 AL5568595.1 AL5569892.1 AL557037.1 AL574092.1 AL574207.1 AL574644.1 AL575373.1 AL577265.1 AW473748.1 BE893192.1 BE896223.1 BE898209.1 BF677721.1 BF689672.1 BF690566.1 BF726069.1 BF998687.1 BG006306.1 BG006608.1 BG282300.1 BG282310.1 BG326634.1 BG422562.1 BG435601.1 BG478666.1 BG676410.1 BG701677.1 BG767953.1 B1040702.1 B1553925.1 B1559648.1 BM016522.1 BM542130.1 BM542378.1 BM549311.1 BM557411.1 BM675641.1</p>	

```

BM906016.1 BM931606.1 BM989082.1 BM993373.1 BM991450.1
BQ18379.1 BQ214877.1 BQ425746.1 BQ574118.1 BQ722286.1
BQ80658.1 BU165852.1 BU171456.1 BU189399.1 CA307415.1
CA414285.1 H51540.1 N26400.1 N26407.1 R88167.1 T48709.1
W22615.1
match: cDNAs: AB075833.1 AK027375.1 AK075381.1 AL8333203.1"
join(complement(50428..50517),complement(39256..39300),
complement(25552..25715),complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:27378..27721))
/genes="ITIH5"
/locus_tag="RP11-67687.1-001"
/standard_name="OTTHUMP00000019062"
/notes="match: proteins: Q8TF48 Q96K70"
/codon_start=1
/product="inter-alpha (globulin) inhibitor H5"
/protein_id="CAI12955.1"
/db_xref="GI:55958060"
/db_xref="GOA:Q5T665"
/db_xref="InterPro:IPR002035"
/db_xref="InterPro:IPR006587"
/db_xref="InterPro:IPR010600"
/db_xref="UniProt/TREMBL:Q5T665"
/translation="MLLLIGLGLSLCVSGSEAEQAQSGHSEODGLRVPQVPELLOR
LKTPLMTEFVKSTIISRYATFTVSCRMNRASEDQDIEFQMOIPIAAFTINTMLI
GDKVYQGEIKREKSGDKVKEKGTTEENGKEGTIFRASAVIPSKDAAPFLSYE
ELLQRRICKYRHSISVRPQQLSGRLSDVNLISAGIASLEVLPHNSRQSGRGED
NGYPPVTVINQNETFANIIFPKTVVOQARIAQNGILGDFIIRYDVNREQSIGDIQVL
DSGFTVHIFAPKDLPLPKNVVFLDSSASMGVTKLROTKDALFILHDLRPQDFR511
GFSNRKIKWKDHLISVTPDSIRDGKVIYHMSPTGGTDINGALQARILRLNKYVAHS
IGDRSVLITVDKPGTGTHTIKLNNTRAEARGQVCIFTIGINDVDVFLLEKL
SLENGCTRRVHEEDAGSLIGFYDEIRLPLSDIRIDYPPSVVQATKTLFNFYF
GSEIIIAKLVDRKLDHLHVEVTASNSKKFIILKTDVPRPQAKGVTSPPRGDG
EGDTNHIIRLWSYLTLLKELLSWQSDDEPEKRLRQALAVSYRFLPTFSMKLR
GPVPRMGLSEAHGMSRAMPPEPVQSVRGAGTOPGLLKKPYQPRIKSITSVDGDP
HFVDFPPLSRITVCENIDGQPDQDILRLVSDHRDGSVTYNGELIGAPAPPNGHKQRTY
Query Match 100.0%; Score 22; DB 8; Length 174361;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAAAGTGATCGCTCCCGTGGT 22
|||||
Db 171584 AGAAAGTGATCGCTCCCGTGGT 171563

RESULT 8
AL391687/c AL391687 194563 bp DNA linear HTG 20-OCT-2001
LOCUS Homo sapiens chromosome 10 clone RP11-299D3, 2 unordered pieces.
DEFINITION AL391687
ACCESSION AL391687
VERSION AL391687.8 GI:14715355
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
1
Almeida, J.
Direct Submission
Submitted (16-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
request: clonerequest@sanger.ac.uk
On Jul 12, 2001 this sequence version replaced gi:13990116.

us-09-555-529-18.rge
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA299D3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 194262 bases at least Q40
Consensus quality: 194328 bases at least Q30
Consensus quality: 194372 bases at least Q20
Insert size: 194463; sum-of-contigs
Insert size: 173753; 41.5% error; agarose-fp
Quality coverage: 13.33x in Q20 bases; sum-of-contigs Quality
coverage: 14.92x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 172067: contig of 172067 bp in length
* 172068 172167: gap of 100 bp
* 172168 194563: contig of 22396 bp in length.
FEATURES
source
1..194563
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone_lib="RPCI-11.2"
1..172067
/notes="assembly fragment:00152
fragment_chain:1"
172168..194563
/notes="assembly fragment:03892
fragment_chain:1
clone_end:T7
vector_side:right"
ORIGIN
Query Match 100.0%; Score 22; DB 14; Length 194563;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAAAGTGATCGCTCCCGTGGT 22
|||||
Db 90730 AGAAAGTGATCGCTCCCGTGGT 90709

RESULT 9
AC017711/c AC017711 7111 bp DNA linear HTG 10-DEC-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
DEFINITION AC017711
ACCESSION AC017711
VERSION AC017711.1 GI:6554286
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 7111)
REFERENCE
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,

```

```

Rockville, MD, USA
This sequence was identified as CDM:10211824 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
    source
        1..7111
            /organism="Drosophila melanogaster"
            /mol_type="genomic DNA"
            /db_xref="taxon:7227"

ORIGIN
Query Match      83.6%; Score 18.4; DB 14; Length 7111;
Best Local Similarity 95.0%; Pred. No. 7.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  2  GAAAGTATCGTCCGCTGG 21
      |||||
Db  6348 GAAAGTATCGCTACCGTGG 6329

RESULT 10
CR847851
LOCUS
DEFINITION
    Danio rerio chromosome 14 clone DKEYP-115D7, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
ACCESSION
    CR847851
VERSION
    CR847851.4 GI:67845775
KEYWORDS
    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
    Danio rerio (zebrafish)
ORGANISM
    Danio rerio
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
    Cypriniformes; Cyprinidae; Danio.
    1 (bases 1 to 166348)
    McLaren, S.
REFERENCE
    Direct Submission
AUTHORS
    Submitted (14-JUN-2005) Wellcome Trust Sanger Institute, Hinxton,
TITLE
    Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
JOURNAL
    zfish-help@sanger.ac.uk Clone requests:
    http://www.sanger.ac.uk/Projects/D/erio/faqs.shtml#dataeight
    On Jun 15, 2005 this sequence version replaced gi:54019839.
    ----- Genome Center
    Center: Wellcome Trust Sanger Institute
    Center code: SC
    Web site: http://www.sanger.ac.uk
    Contact: zfish-help@sanger.ac.uk
    ----- Project Information
    Center project name: zkp115D7
    ----- Summary Statistics
    Assembly program: XGAP4; version 4.5
    Chemistry: Dye-terminator; 100% of reads
    Consensus quality: 161693 bases at least Q40
    Consensus quality: 162463 bases at least Q30
    Consensus quality: 163159 bases at least Q20
    Insert size: 165448; sum-of-contigs
    Insert size: 191089; 11.8% error; agarose-fp
    Quality coverage: 8.42x in Q20 bases; sum-of-contigs Quality
    coverage: 7.39x in Q20 bases; agarose-fp
    -----
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 10 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
    * arbitrary. Gaps between the contigs are represented as
    * runs of N, but the exact sizes of the gaps are unknown.
    * This record will be updated with the finished sequence
    * as soon as it is available and the accession number will
    * be preserved.
    *
    * 1 4408: contig of 4408 bp in length
    * 4509 4508: gap of 100 bp
    * 4509 15390: contig of 10882 bp in length

```

```

* 15391 15490: gap of 100 bp
* 15491 30342: contig of 14852 bp in length
* 30343 30442: gap of 100 bp
* 30443 77547: contig of 47105 bp in length
* 77548 77647: gap of 100 bp
* 77648 113849: contig of 36202 bp in length
* 113850 113949: gap of 100 bp
* 113950 119804: contig of 5855 bp in length
* 119805 119904: gap of 100 bp
* 119905 124866: contig of 4962 bp in length
* 124867 124967: gap of 100 bp
* 124968 135191: contig of 10225 bp in length
* 135192 135292: gap of 100 bp
* 135293 149712: contig of 14321 bp in length
* 149713 149713: gap of 100 bp
* 149713 166348: contig of 16636 bp in length.

FEATURES
    Location/Qualifiers
        1..166348
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7955"
            /chromosome="14"
            /clone="DKEYP-115D7"
            /clone_lib="DanioKeyPilot"
        1..4408
            /notes="assembly_fragment:00030"
            fragment_chain:1
        4509..15390
            /notes="assembly_fragment:00274"
            fragment_chain:1
        15491..30342
            /notes="assembly_fragment:00615"
            fragment_chain:1
        30443..77547
            /notes="assembly_fragment:01921"
            fragment_chain:2
        77648..113849
            /notes="assembly_fragment:01377"
            fragment_chain:2
        113950..119804
            /notes="assembly_fragment:00183"
            fragment_chain:2
        119905..124866
            /notes="assembly_fragment:00099.0"
            fragment_chain:2
        124967..135191
            /notes="assembly_fragment:00425"
            fragment_chain:2
        135292..149612
            /notes="assembly_fragment:00835"
            fragment_chain:2
        149713..166348
            /notes="assembly_fragment:01070"

ORIGIN
Query Match      83.6%; Score 18.4; DB 14; Length 166348;
Best Local Similarity 95.0%; Pred. No. 6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy  1  AGAAAGTATCGTCCGCTG 20
      |||||
Db  97301 AGAAAGTATCGTCCATG 97320

RESULT 11
AC146480/c
LOCUS
DEFINITION
    Danio rerio clone CH211-4009, complete sequence.
ACCESSION
    AC146480
VERSION
    AC146480.2 GI:37951362
KEYWORDS
    HTG.
SOURCE
    Danio rerio (zebrafish)
ORGANISM
    Danio rerio
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
    Cypriniformes; Cyprinidae; Danio.

```

REFERENCE
AUTHORS Talbot,W.S., Rauch,G.J., Noonan,J., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
TITLE Genomic studies of vertebrate diversity
JOURNAL Unpublished
REFERENCE
AUTHORS Talbot,W.S., Rauch,G.J., Noonan,J., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-2003) Stanford Human Genome Center, 975 California Avenue, Palo Alto, CA 94304, USA
REFERENCE
AUTHORS Talbot,W.S., Rauch,G.J., Noonan,J., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
TITLE Direct Submission
JOURNAL Submitted (24-OCT-2003) Stanford Human Genome Center, 975 California Avenue, Palo Alto, CA 94304, USA
COMMENT On Oct 24, 2003 this sequence version replaced gi:33859862. The sequence of the clone was established as a mapping and sequencing collaboration at the Stanford Genome Evolution Center, funded by the NIH Centers of Excellence in Genomic Science (CEGS) initiative (http://cegs.stanford.edu). The clone was isolated from the BAC library CHORI-211 (http://bacpac.chori.org).

Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.

FEATURES
source
Location/Qualifiers
1..166906
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-4009"

ORIGIN
Query Match 83.6%; Score 18.4; DB 5; Length 166906;
Best Local Similarity 95.0%; Pred. No. 6e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AGAAAGTGCCTGCTGCGCTG 20
|||||
Db 146447 AGAAAGTGCCTGCTGCGCATG 146428

RESULT 12
AL929558/c
LOCUS AL929558 170597 bp DNA linear VRT 01-NOV-2003
DEFINITION Zebrafish DNA sequence from clone CH211-159N7, complete sequence.
ACCESSION AL929558
VERSION AL929558.13 GI:38143466
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE
AUTHORS 1 (bases 1 to 170597)
Whitehead,S.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk
COMMENT On Nov 1, 2003 this sequence version replaced gi:37718665. ----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.
This sequence was finished as follows otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep
Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml CH211-159N7 is from a CHORI-211 BAC library

VECTOR: pTARBAC2.1.

FEATURES
source
Location/Qualifiers
1..170597
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-159N7"
/clone_lib="CHORI-211"

ORIGIN

Query Match 83.6%; Score 18.4; DB 5; Length 170597;
Best Local Similarity 95.0%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AGAAAGTGCCTGCTGCGCTG 20
|||||

Db 146404 AGAAAGTGCCTGCTGCGCATG 146385

RESULT 13

AC018491
LOCUS AC018491 177953 bp DNA linear INV 31-JUL-2004
DEFINITION Drosophila melanogaster clone BACR11H15, complete sequence.
ACCESSION AC018491
VERSION AC018491.9 GI:50872279
KEYWORDS HTG.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS 1 (bases 1 to 177953)

Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Buehner,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Richards,S., Sethi,H., Svirskaas,R.R., Wan,K.H., Webster,D., Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.

Direct Submission

TITLE Submitted (13-DEC-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

JOURNAL 2 (bases 1 to 177953)

REFERENCE
AUTHORS Celniker,S., Carlson,J., Wan,K., Pfeiffer,B., Frisoe,E., George,R., Hoskins,R., Stapleton,M., Pacieb,J., Park,S., Svirskaas,R., Smith,E., Yu,C. and Rubin,G.

Direct Submission

TITLE Submitted (31-JUL-2004) Berkeley Drosophila Genome Project, MS

JOURNAL

64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, US
On Jul 31, 2004 this sequence version replaced gi:14701992.
Sequence submitted by:

Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

This sequence submission incorporates changes made during reevaluation of the assembly or fingerprint verification of the clone. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.org.

FEATURES

source

1. 177953
Location/Qualifiers
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="X"
/map="19A-198"
/clone="BAC clone BACR11H15 (D1225)"
/clone_lib="RPC1-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBACs.6)"

ORIGIN

Query Match 83.6%; Score 18.4; DB 2; Length 177953;
Best Local Similarity 95.08; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 GAAAGTGATCGTGCCTGG 21

Db 146451 GAAAGTGATCGTACCCTGG 146470

RESULT 14

AB002611 297848 bp DNA linear INV 10-AUG-2005
LOCUS Drosophila melanogaster chromosome X, section 66 of 74 of the complete sequence.

AB002611 AE014298

AB002611.5 GI:55380356

Drosophila melanogaster (fruit fly)

Drosophila melanogaster

Eukaryota: Metazoa: Arthropoda: Insecta: Pterygota;

Neoptera: Endopterygota; Diptera: Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 297848)

Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Ananidis,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brannon,R.C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D., Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor,G.L., Abail,J.F., Agbayani,A., An,H.J., Andrews-Pfankuch,C., Baldwin,D., Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D., Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S., Fleischmann,W., Fosler,C., Gabriellian,A.E., Garg,N.S., Geibart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Laoko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattel,B., McIntosh,T.C.,

McLeod,M.P., McPherson,D., Mervulov,G., Milshina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nuskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,P., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirekas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstein,G.M., Weissenbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
10731132
2 (bases 1 to 297848)
Celniker,S.E., Wheeler,D.A., Kronmiller,B., Carlson,J.W., Halpern,A., Patel,S., Adams,M., Champe,M., Dugan,S.P., Frise,E., Hodgson,A., George,R.A., Hoskins,R.A., Laverty,T., Muzny,D.M., Nelson,C.R., Pacleb,J.M., Park,S., Pfeiffer,B.D., Richards,S., Sodergren,E.J., Svirekas,R., Tabor,P.E., Wan,K., Stapleton,M., Sutton,G.G., Venter,C., Weinstein,G., Scherer,S.E., Myers,E.W., Gibbs,R.A. and Rubin,G.M.
Finishing a whole-genome shotgun: release 3 of the Drosophila melanogaster euchromatic genome sequence
Genome Biol. 3 (12), RESEARCH0079 (2002)
12537568
3 (bases 1 to 297848)
Mitra,S., Crosby,M.A., Mungall,C.J., Matthews,B.B., Campbell,K.S., Hradecky,P., Huang,Y., Kaminker,J.S., Milburn,G.H., Prochnik,S.E., Smith,C.D., Tupy,J.L., Whitfield,E.J., Bayraktaroglu,L., Berman,B.P., Bettencourt,B.R., Celniker,S.E., de Grey,A.D., Drysdale,R.A., Harris,N.L., Richter,J., Russo,S., Schroeder,A.J., Shu,S.Q., Stapleton,M., Yamada,C., Ashburner,M., Gelbart,W.M., Rubin,G.M. and Lewis,S.E.
Annotation of the Drosophila melanogaster euchromatic genome: a systematic review
Genome Biol. 3 (12), RESEARCH0083 (2002)
12537572
4 (bases 1 to 297848)
Kaminker,J.S., Bergman,C.M., Kronmiller,B., Carlson,J., Svirekas,R., Patel,S., Frise,E., Wheeler,D.A., Lewis,S.E., Rubin,G.M., Ashburner,M. and Celniker,S.E.
The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective
Genome Biol. 3 (12), RESEARCH0084 (2002)
12537573
5 (bases 1 to 297848)
Celniker,S., Carlson,J., Wan,K., Pfeiffer,B., Frise,E., George,R., Hoskins,R., Stapleton,M., Pacleb,J., Park,S., Svirekas,R., Smith,E., Yu,C. and Rubin,G.
Berkeley Drosophila Genome Project
Drosophila melanogaster release 4 sequence
Unpublished
6 (bases 1 to 297848)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
7 (bases 1 to 297848)
FlyBase
Direct Submission
Submitted (22-JUL-2005) FlyBase, Harvard University, Biological Laboratories, 16 Divinity Ave, Cambridge, MA 02138, USA
On Nov 4, 2004 this sequence version replaced gi:26381542.
Location/Qualifiers
1. 297848
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

source

```
gene      /chromosome="X"
          complement(<3406..>3833)
          /gene="CG32821"
          /locus_tag="Dmel_CG32821"
          /old_locus_tag="CG32821"
          /map="19A2-19A2"
          /db_xref="FLYBASE:FBgn0052821"
          complement(join(<3406..3623,3722..>3833))
          /gene="CG32821"
          /locus_tag="Dmel_CG32821"
          /old_locus_tag="CG32821"
          /product="CG32821-RA"
          /db_xref="FLYBASE:FBgn0052821"
          complement(join(3406..3623,3722..3833))
          /gene="CG32821"
          /locus_tag="Dmel_CG32821"
          /old_locus_tag="CG32821"
          /note="CG32821 gene product from transcript CG32821-RA"
          /codon_start=1
          /product="CG32821-PA"
          /protein_id="AAN08990.1"
          /db_xref="GI:22831369"
          /db_xref="FLYBASE:FBgn0052821"
          /translation="MEISNAHPNSLARTLTGTAANAANKTOTTTPSLPRSAGNGV
          YVSTDEHGRPFYNNQVHPGQVMPNPGTYQNGCAIIGSSVGVTVTIALGL
          ALGGRP"
          complement(3701..4491)
          /transposon="Jockey"
          /db_xref="FLYBASE:FBti0063732"
          4516..5690
          /map="19A2-19A2"
          /transposon="Rt1c"
          /db_xref="FLYBASE:FBti0019647"
          complement(8865..9655)
          /transposon="Jockey"
          /db_xref="FLYBASE:FBti0063734"
          9680..10854
          /transposon="Rt1c"
          /db_xref="FLYBASE:FBti0062858"
          complement(14175..14462)
          /transposon="Stalker4"
          /db_xref="FLYBASE:FBti0062865"
          complement(14601..14702)
          /transposon="invader2"
          /db_xref="FLYBASE:FBti0062837"
          14703..18743
          /transposon="F"
          /db_xref="FLYBASE:FBti0062825"
          complement(18785..19116)
          /transposon="mdg3"
          /db_xref="FLYBASE:FBti0062826"
          complement(19151..20114)
          /transposon="invader3"
          /db_xref="FLYBASE:FBti0062839"
          complement(19194..19434)
          /transposon="mdg3"
          /db_xref="FLYBASE:FBti0062904"
          complement(22760..23550)
          /transposon="Jockey"
          /db_xref="FLYBASE:FBti0063735"
          23575..24749
          /transposon="Rt1c"
          /db_xref="FLYBASE:FBti0062860"
          complement(28032..28319)
          /transposon="Stalker4"
          /db_xref="FLYBASE:FBti0062867"
          complement(28462..29249)
          /transposon="mdg3"
          /db_xref="FLYBASE:FBti0062828"
          complement(29422..29930)
          /transposon="invader3"
          /db_xref="FLYBASE:FBti0062915"
          complement(32576..33366)
```

```
repeat_region repeat_region
              /transposon="Jockey"
              /db_xref="FLYBASE:FBti0062888"
              33391..34565
              /transposon="Rt1c"
              /db_xref="FLYBASE:FBti0062927"
              complement(37884..38171)
              /transposon="Stalker4"
              /db_xref="FLYBASE:FBti0062940"
              complement(38308..39785)
              /transposon="invader3"
              /db_xref="FLYBASE:FBti0062917"
              complement(42614..42901)
              /transposon="Stalker4"
              /db_xref="FLYBASE:FBti0062942"
              complement(42980..44518)
              /transposon="invader3"
              /db_xref="FLYBASE:FBti0062919"
              complement(47349..47636)
              /transposon="Stalker4"
              /transposon="Stalker4"
              83.6%; Score 18.4; DB 2; Length 297848;
              Query Match
              Best Local Similarity 95.0%; Pred. No. 5.8e+02;
              Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GAAAGTGATCGCTCCCGTGG 21
    |||||
Db 237313 GAAAGTGATCGCTACCGTGG 237332

RESULT 15
LOCUS AY525807
DEFINITION Streptococcus pneumoniae strain HK_P116 Gdh (gdh) gene, partial cds.
ACCESSION AY525807
VERSION AY525807.1 GI:46486224
KEYWORDS
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 460)
AUTHORS Ko, K.S., Oh, W.S., Peck, K.R., Lee, J.H., Lee, N.Y. and Song, J.H.
TITLE Phenotypic and genotypic discrepancy of Streptococcus pneumoniae strains isolated from Asian countries
JOURNAL FEMS Immunol. Med. Microbiol. 45 (1), 63-70 (2005)
PUBMED 15985224
REFERENCE 2 (bases 1 to 460)
AUTHORS Ko, K.S. and Song, J.-H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2004) Asian-Pacific Research Foundation for Infectious Diseases (ARFID), 50, Ilwon-dong, Kangnam-ku, Seoul 135-710, Korea
FEATURES
source
    1..460
    /organism="Streptococcus pneumoniae"
    /mol_type="genomic DNA"
    /strain="HK_P116"
    /db_xref="taxon:1313"
    <1..>460
    /gene="gdh"
    <1..>460
    /gene="gdh"
    /codon_start=2
    /transl_table=11
    /product="Gdh"
    /protein_id="AAS98629.1"
    /db_xref="GI:46486225"
    /translation="EHPFRGQYRSKIDGMKYISVRSEPNVNPSTTTTFASGAPFVD
    SDRFRGVFFPRTGKRTKTHVNVFQKQMSIFGRLPLANILITIIQPTGFSLSL
    NGKQVGEFNLAPNSLDYRTDATATGASPEPYEKLIIYDLVNNNSTNFSWD"

ORIGIN
```

Query Match 80.9%; Score 17.8; DB 1; Length 460;
Best Local Similarity 90.5%; Pred.No. 1.6e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAAAGTCATCGCTGCCCGTGGT 22
|||
Db 131 GACAGTCATCGCTGCCCGTGGT 151
|||

Search completed: November 26, 2005, 19:11:04
Job time : 121.932 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100	0	26	6	AX003326 Sequence
c 2	26	100	0	616	10	BV448133 KIN 9291
c 3	26	100	0	1002	6	AX003310 Sequence
c 4	26	100	0	1528	8	AX003308 Sequence
c 5	26	100	0	1528	8	HSAJ5273 Homo sapi
c 6	26	100	0	2069	8	BC017309 Homo sapi
c 7	26	100	0	174361	8	AL158044 Human DNA
c 8	26	100	0	194563	14	AL391687 Homo sapi
c 9	20.2	77.7	110000	2	CP000079.01	Continuation (2 of
c 10	19.8	76.2	1102	6	AX003303 Sequence	
c 11	19.8	76.2	1390	6	AX003331 Sequence	
c 12	19.8	76.2	1414	9	MMK1N17	
c 13	19.8	76.2	2710	1	BC059169 Mus muscu	
c 14	19.8	76.2	19113	1	CR931708 Streptoco	
c 15	19.8	76.2	86263	14	AF0061973 Lotusc or	
c 16	19.8	76.2	144118	14	AC141995 Rattus no	
c 17	19.8	76.2	180882	14	AC144615 Mus muscu	
c 18	19.8	76.2	182400	14	AC124010 Mus muscu	

KEYWORDS
STS
Macaca mulatta (rhesus monkey)

SOURCE
Macaca mulatta

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Cercopithecoidea; Cercopitheciinae; Macaca.

REFERENCE
1 (bases 1 to 616)
Spindel, E.R., Pauley, M., Jia, Y., Thompson, S., Lanthorn, M., Gravett, C., Lupo, S.L., Tchourbanov, A., Ali, H., Ojeda, S.R., and Norgren, R. B.

AUTHORS
Targeted amplification of the 3' end of rhesus macaque orthologs of human genes
Unpublished (2004)

TITLE
Contact: Spindel, ER
Division of Neuroscience
Oregon National Primate Research Center
505 NW 185th Avenue, Beaverton, OR 97006, USA
Tel: 403-690-5388
Fax: 503-690-5384
Email: spindel@ohsu.edu
Primer A: ggaattcaatcagaacatc
Primer B: tacattgctgcgtctcag
STS size: 616

JOURNAL
COMMENT

PCR Profile:
Hot Start: 95 degrees C for 2.00 min
Denaturation: 95 degrees C for 0.50 min
Annealing: 51 degrees C for 0.50 min
Polymerization: 72 degrees C for 1.00 min
PCR Cycles: 35
Extension: 72 degrees C for 7.0 min
Thermal Cycler: MJ Instruments PTC100

Protocol:
Template: 200 ng
Primer: each 1uM
dNTP's: each 200 uM
Tag Polymerase: 0.05 units/uL (Fast Start High Fidelity, Roche)
Total Vol: 50 uL

Buffer:
MgCl2: 1.8 mM
Fast Start polymerase reaction buffer (Roche)

FEATURES
source
1..616
Location/Qualifiers
/organism="Macaca mulatta"
/mol_type="genomic DNA"
/strain="Indian origin"
/db_xref="taxon:9544"
/clone="MMA9291"
/clone_1b="Rhesus macaque genomic DNA"
/dev_stage="Adult"
/note="Organ: Liver; Vector: pGEM-T Easy; V-type: Plasmid; STS was amplified from rhesus genomic DNA with the human forward and reverse primers listed above and subcloned into pGEM-T Easy"
1..616
/gene="KIN"
/note="KIN, antigenic determinant of rca protein homolog (mouse)"
STS
<1..>616

ORIGIN
/gene="KIN"

Query Match
Best Local Similarity 100.0%; Score 26; DB 10; Length 616;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 GCGACACCAATTTGATGCTTTAGA 26
71 GCGACACCAATTTGATGCTTTAGA 46

RESULT 3
AX003310/c
LOCUS
DEFINITION
Sequence 3 from Patent WO929845.
AX003310
ACCESSION
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo;

REFERENCE
AUTHORS
Angulo-Mora, J.F. and Mauffrey, P.
TITLE
Sequences coding for kin17 protein and their applications
JOURNAL
Patent: WO 929845-A 3 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)

FEATURES
source
1..1002
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 1002;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 GCGACACCAATTTGATGCTTTAGA 26
1002 GCGACACCAATTTGATGCTTTAGA 977

RESULT 4
AX003308/c
LOCUS
DEFINITION
Sequence 1 from Patent WO929845.
AX003308
ACCESSION
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
AUTHORS
Angulo-Mora, J.F. and Mauffrey, P.
TITLE
Sequences coding for kin17 protein and their applications
JOURNAL
Patent: WO 929845-A 1 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)

FEATURES
source
1..1528
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match
Best Local Similarity 100.0%; Score 26; DB 6; Length 1528;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Wed Nov 30 09:40:42 2005

us-09-555-529-19.rge

atches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GCGAACACCAATTGATGCTTTAAGA 26

1303 GCGAACACCAATTGATGCTTTAAGA 1278

RESULT 5
HSAJ5273/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

HSAJ5273
Homo sapiens mRNA for Kin17 protein.
AJ005273.1 GI:3850703
Kin17 gene.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
AUTHORS
TITLE
JOURNAL
REMARK
AUTHORS
TITLE
JOURNAL
COMMENT

1. 1528
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/sex="male"
/tissue_type="testis"
1. 1528
/gene="Kin17"
66. 1247
/gene="Kin17"
/codon_start=1
/protein_id="CAA06462.1"
/db_xref="GI:3850704"
/db_xref="GGA:060870"
/db_xref="UniProt/TREMBL:O60870"
Translation="MGKSPDLPFKAIANRKSGLQKLRWYCOMCKOCRDENGFKCH
CMSEIQORALLASNPQDPSEFRNDFLELRPRGTRVHNNIVVYVYISHR
EHIMQATQETLDTFKMLRGELCKVDPTPKGWYIQYIDRPETIPROLEKCKK
QDLDDEKTAKEFIEQVRGLGKQEVPTFELSDRENDEKVTFLNKGACSSGAT
SSKSTTIPGALKTISSASVKRKSQSSQSKRKKKKAALDEIMETIEEKKTAR
TDYWLQPEITVITKTKGKKKAIKVEIDKYTAVVKNLDSGDKLKDQTHLET
VIPAPGRKILVNGRGYRNEGTLESINEKTSATIVTETPLKGRVEGIQYEDISK
A"

Query Match 100.0%; Score 26; DB 8; Length 1528;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGAACACCAATTGATGCTTTAAGA 26

Db 1303 GCGAACACCAATTGATGCTTTAAGA 1278

RESULT 6
BC017309/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BC017309
Homo sapiens KIN, antigenic determinant of recA protein homolog
(mouse). mRNA (CDNA clone MGC:29595 IMAGE:5089578), complete cds.
BC017309
BC017309.2 GI:40226034
MGC.
Homo sapiens (human)
Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT

1 (bases 1 to 2069)
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.P., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
Diatchenko, L., Marusik, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schetz, T.E., Brownstein, M.J., Udell, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McSwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Sahney, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywicki, M.I., Skalska, U., Smalish, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932

2 (bases 1 to 2069)
Straussberg, R.
Direct Submission
Submitted (05-NOV-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Dec 19, 2003 this sequence version replaced gi:16878214.
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nslc.nih.gov/>
Contact: nslc@ncntrgri.nih.gov
Akter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granice, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-t., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripp, S., Thomas, P.J., Touchman, J.W.,
Turgowell, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, B.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IMAGE Plate: 39 Row: 0 Column: 18
This clone was selected for full length sequencing because it

FEATURES
Source

passed the following selection criteria: matched mRNA gi: 13124882.

Location/Qualifiers

1. 2069
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:29595 IMAGE:5089578"
/cbase_type="Pancreas, epithelioid carcinoma"
/clone_1lb="NH_MGC_42"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1. 2069
/gene="KIN"
/note="synonym: KIN17, BTCD"
/db_xref="GeneID:22944"
/db_xref="MIM:601720"
12. 1193
/gene="KIN"
/codon_start=1
/product="Hsk17 protein"
/protein_id="AAH17309.1"
/db_xref="GI:16878215"
/db_xref="GeneID:22944"
/db_xref="MIM:601720"
/translation="MGSDPLTPKAIANKIKGQKQKRWYCOMQCKQCDENGFKCH
CHSHQKQILASNPQPMDFSEFRNDLELRRTKRVKNNIVNYEYISHR
EIHNNATQWETLTDPTKMLRGGLCKVDPTKGYIYIDRPTIRQLLEKKK
QDLDEKTAKEIEQVRGIEGKEQVPTFELSRNDEKVTFLNLSKACSSGAT
SSKSTSLPSALKTIGSSAKRKSOSSTQSKKKKSKALDEINLEEEKKRRAT
TVMLOPPIVITIKLGEKHKKAIKEVIDYTAVMVMDGKMLDQTHLET
VTPAGKRILVNGVGRNGEGLSINEKTPSATIVITGPIKGRVREGIYQYEDISK
A"

gene

CDS

ORIGIN

Query Match 100.0%; Score 26; DB 8; Length 2069;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGAGCCCAATTTGATGCTTTAAGA 26
Db 1249 GCGAGCCCAATTTGATGCTTTAAGA 1224

RESULT 7
AL158044 LOCUS 174361 bp DNA linear PRI 18-MAY-2005
DEFINITION Human DNA sequence from clone RP11-264C14 on chromosome 10 contains
the 5' end of the gene for a novel protein (MGC10848), the ITH2
gene for inter-alpha (globulin) inhibitor, H2 polypeptide (H2P),
the KIN gene for KIN, antigenic determinant of reca protein homolog
(mouse), the 3' end of the ATP5C1 gene for ATP synthase, H+
transporting, mitochondrial F1 complex, gamma polypeptide 1,
(ATP5C1, ATP5C1), a novel gene and two CpG islands, complete
sequence.

AL158044
AL158044.17 GI:16214543
HTG; ATP5C1; ATP5C1; CpG island; H2P; ITH2; KIN; MGC10848.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo
1 (bases 1 to 174361)
Direct Submision
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
On Oct 17, 2001 this sequence version replaced gi:15131961.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at

FEATURES

Source

http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-264C14 is from the library RP11-11.1 constructed by the group
of Peter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.

misc_feature

gene

mRNA

CDS

1. 174361
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-264C14"
/clone_1lb="RP11-11.1"
/note="Clone left end: RP11-264C14"
join(complement(3206..3285).
complement(AL355374.14:26760..81619))
/gene="ITH5"
/locus_tag="RP11-676E7.1-003"
join(complement(3206..3285).
complement(AL355374.14:81450..81619).
complement(AL355374.14:80319..80435).
complement(AL355374.14:50238..50406).
complement(AL355374.14:44092..44401).
complement(AL355374.14:40790..41349).
complement(AL355374.14:37479..37532).
complement(AL355374.14:34005..34121).
complement(AL355374.14:30367..30744).
complement(AL355374.14:26760..27721))
/gene="ITH5"
/locus_tag="RP11-676E7.1-003"
/product="inter-alpha (globulin) inhibitor H5"
/note="match: ESTs: BE386984.1 BE387282.1 BG33950.1
BG46698.1 B0225649.1
match: CDNA: BC004282.1"
join(complement(3206..3215).
complement(AL355374.14:81450..81619).
complement(AL355374.14:80319..80435).
complement(AL355374.14:50238..50406).
complement(AL355374.14:44092..44401).
complement(AL355374.14:40790..41349).
complement(AL355374.14:37479..37532).
complement(AL355374.14:34005..34121).
complement(AL355374.14:30367..30744).
complement(AL355374.14:27378..27721))
/gene="ITH5"
/locus_tag="RP11-676E7.1-003"
/standard_name="OTHOMPO000019064"
/note="match: proteins: Q9BR94"
/codon_start=1
/product="inter-alpha (globulin) inhibitor H5"
/protein_id="CAI12953.1"
/db_xref="GI:55958058"
/db_xref="GCA:Q5T664"
/db_xref="InterPro:IPR002035"
/db_xref="InterPro:IPR010600"

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:41:08 ; Search time 7.76471 Seconds
(without alignments)
421.056 Million cell updates/sec

Title: US-09-555-529-18
Perfect score: 22
Sequence: 1 agaaagtgcgtcgctgggt 22

Scoring table: IDENTITY_INUC
Gapop 1010 , Gapext 1.0

Searched: 3205263 seqs, 74304013 residues

Total number of hits satisfying chosen parameters: 6410526

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA New:*

1: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	15.2	69.1	333	1	US-10-802-796-723
C 2	14.6	66.4	163	1	US-10-802-796-492
C 3	14.6	66.4	262	1	US-10-802-796-275
C 4	14.6	66.4	264	1	US-10-802-796-228
C 5	14.6	66.4	295	1	US-10-802-796-448
C 6	14.6	66.4	325	1	US-10-802-796-216
C 7	14.6	66.4	420	1	US-10-802-796-496
C 8	14.6	66.4	428	9	US-11-082-389-413
C 9	14.6	66.4	444	1	US-10-821-234-418
C 10	14.6	66.4	445	1	US-10-802-796-75
C 11	14.6	66.4	470	1	US-10-802-796-157
C 12	14.6	66.4	1575	1	US-10-821-234-779
C 13	14.2	64.5	19	8	US-11-101-244-16138
C 14	14.2	64.5	19	9	US-11-083-784-16138
C 15	14.2	64.5	1018	1	US-10-821-234-225
C 16	14	63.6	1277	1	US-10-821-234-9
C 17	14	63.6	1511	7	US-11-102-978-6
C 18	13.8	62.7	1509	9	US-11-082-389-411
C 19	13.8	62.7	3341	7	US-11-110-837-1
C 20	13.8	62.7	3341	7	US-11-110-837-3
C 21	13.6	61.8	247	1	US-10-802-796-475
C 22	13.6	61.8	889	9	US-11-082-389-119
C 23	13.6	61.8	957	9	US-11-082-389-117

24	13.6	61.8	1343	9	US-11-082-389-105	Sequence 105, App
25	13.6	61.8	1575	1	US-10-821-234-201	Sequence 201, App
26	13.6	61.8	1830	9	US-11-082-389-103	Sequence 103, App
27	13.6	61.8	2058	7	US-11-065-943-57	Sequence 57, Appl
C 28	13.6	61.8	2163	1	US-10-131-826A-127	Sequence 127, App
C 29	13.6	61.8	3449	1	US-10-131-826A-293	Sequence 293, App
C 30	13.6	61.8	3684	1	US-10-858-730-261	Sequence 261, App
31	13.4	60.9	19	8	US-11-101-244-13019	Sequence 213019, App
32	13.4	60.9	19	8	US-11-101-244-1021672	Sequence 1021672, App
C 33	13.4	60.9	19	8	US-11-101-244-1035689	Sequence 1035689, App
C 34	13.4	60.9	19	9	US-11-083-784-213019	Sequence 213019, App
C 35	13.4	60.9	19	9	US-11-083-784-1021672	Sequence 1021672, App
C 36	13.4	60.9	19	9	US-11-083-784-1035689	Sequence 1035689, App
C 37	13.2	60.0	19	8	US-11-101-244-139395	Sequence 139395, App
C 38	13.2	60.0	19	8	US-11-101-244-452828	Sequence 452828, App
C 39	13.2	60.0	19	8	US-11-101-244-1250329	Sequence 1250329, App
C 40	13.2	60.0	19	8	US-11-101-244-1449736	Sequence 1449736, App
C 41	13.2	60.0	19	8	US-11-101-244-1449781	Sequence 1449781, App
C 42	13.2	60.0	19	9	US-11-083-784-139395	Sequence 139395, App
C 43	13.2	60.0	19	9	US-11-083-784-452828	Sequence 452828, App
C 44	13.2	60.0	19	9	US-11-083-784-1250329	Sequence 1250329, App
C 45	13.2	60.0	19	9	US-11-083-784-1449736	Sequence 1449736, App

ALIGNMENTS

RESULT 1
US-10-802-796-723/c
; Sequence 723, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 723
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (45)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (64)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (130)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (146)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base

```

; LOCATION: (205)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (211)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (271)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (309)
; OTHER INFORMATION: a, t, c or g
US-10-802-796-723

```

Query Match 69.1%; Score 15.2; DB 1; Length 333;

Best Local Similarity 81.0%; Pred. No. 57; Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 2 GAAAGTGATCGCTGCCGTGGT 22
    ||| ||| ||| ||| ||| ||| |||
Db 133 GAANGTGGCTGCTGGCGTCT 113

```

RESULT 2

```

US-10-802-796-492/c
; Sequence 492, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 492
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-802-796-492

```

Query Match 66.4%; Score 14.6; DB 1; Length 163;

Best Local Similarity 81.0%; Pred. No. 1.1e+02; Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 2 GAAAGTGATCGCTGCCGTGGT 22
    ||| ||| ||| ||| ||| ||| |||
Db 103 GAATGTGGCTGCTGGCGTCT 83

```

RESULT 3

```

US-10-802-796-275/c
; Sequence 275, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN

```

```

; LOCATION: (205)
; OTHER INFORMATION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 275
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (21)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (39)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (97)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (122)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (130)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (144)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (222)..(223)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (225)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (229)
; OTHER INFORMATION: a, t, c or g
US-10-802-796-275

```

Query Match 66.4%; Score 14.6; DB 1; Length 262;

Best Local Similarity 81.0%; Pred. No. 1.1e+02; Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 2 GAAAGTGATCGCTGCCGTGGT 22
    ||| ||| ||| ||| ||| ||| |||
Db 66 GAATGTGGCTGCTGGCGTCT 46

```

RESULT 4

```

US-10-802-796-228/c
; Sequence 228, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND

```



```
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (21)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (23)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (106)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (200)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (272)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (355)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (362)
; OTHER INFORMATION: a, t, c or g
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (413)
; OTHER INFORMATION: a, t, c or g
; OTHER INFORMATION: a, t, c or g
US-10-802-796-496
```

```
Query Match 66.4%; Score 14.6; DB 1; Length 420;
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 GAAAGTGATCGCTGCCGTGGT 22
Db 128 GAATGTGGTGGCTGGCGTGGT 108
|||||
```

```
RESULT 8
US-11-082-389-413/c
; Sequence 413, Application US/11082389
; Publication No. US20050244935A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORVNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
; TITLE OF INVENTION: TRANSPORT
; FILE REFERENCE: BGI-131CPCN
; CURRENT APPLICATION NUMBER: US/11/082,389
; PRIOR FILING DATE: 2005-03-16
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 09/603024
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 60/143262
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: US 60/151281
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: DE 19930487.4
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19930489.0
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19931549.3
; PRIOR FILING DATE: 1999-07-08
```

```
; PRIOR APPLICATION NUMBER: DE 19931550.7
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932134.5
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19941379.7
; PRIOR FILING DATE: 1999-08-31
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 446
; SEQ ID NO 413
; LENGTH: 428
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(405)
; OTHER INFORMATION: RXA02628
US-11-082-389-413
```

```
Query Match 66.4%; Score 14.6; DB 9; Length 428;
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 GAAAGTGATCGCTGCCGTGGT 22
Db 237 GAAGTTGATCGCTGGGTGAT 217
|||||
```

```
RESULT 9
US-10-821-234-418
; Sequence 418, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preseclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 418
; LENGTH: 444
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-418
```

```
Query Match 66.4%; Score 14.6; DB 1; Length 444;
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 AGAAAGTGATCGCTGCCGTGG 21
Db 395 AGAAAGTGCTGGCTGGTGG 415
|||||
```

```
RESULT 10
US-10-802-796-75/c
; Sequence 75, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
```



```
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 16138
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-16138
```

```
Query Match      64.5%; Score 14.2; DB 9; Length 19;
Best Local Similarity 63.2%; Pred. No. 1.4e+02;
Matches 12; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2 GAAAGTCATCGCTGCCGTG 20
         |||||:|:|:|:|:|
Db       1 GAAAGUGAUGCGGCUAUG 19
```

```
RESULT 15
US-10-821-234-225
; Sequence 225, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 225
; LENGTH: 1018
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-225
```

```
Query Match      64.5%; Score 14.2; DB 1; Length 1018;
Best Local Similarity 84.2%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      3 AAAGTCATCGCTGCCGTGG 21
         ||||| | ||||| | ||
Db       204 AAAGTCACAGCTGCCATGG 222
```

Search completed: November 27, 2005, 02:44:30
Job time : 7.76471 secs

RESULT 14
US-11-083-784-16138
; Sequence 16138, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 : Search time 25.3623 Seconds
(without alignments)
7173.088 Million cell updates/sec

Title: US-09-555-529-18

Perfect score: 22

Sequence: 1 agaaagtgcgcgtgcgtggt 22

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	22	100.0	588	4	US-09-925-065A-20834, A
C 2	17.4	79.1	381	3	US-09-732-627A-1180, Ap
C 3	17.4	79.1	1946	9	US-10-450-763-21563, A
C 4	17.4	79.1	3031	7	US-10-655-042-13, Appl
C 5	17.4	78.1	3307	7	US-10-655-042-15, Appl
C 6	17.4	79.1	3334	7	US-10-655-042-5, Appl
C 7	17.4	79.1	3388	7	US-10-655-042-7, Appl
C 8	17.4	79.1	3487	7	US-10-655-042-11, Appl
C 9	17.4	79.1	3532	7	US-10-655-042-3, Appl
C 10	17.4	79.1	3586	7	US-10-655-042-9, Appl
C 11	17.4	78.1	3661	7	US-10-633-577-1, Appl
C 12	17.4	79.1	3661	7	US-10-655-042-1, Appl
C 13	17.2	78.2	359	8	US-10-425-115-17122, A
C 14	17.2	78.2	412	3	US-09-864-761-21719, A
C 15	17.2	78.2	421	3	US-09-864-761-4990, A
C 16	17.2	78.2	475	7	US-10-296-115-58, Appl
C 17	17.2	78.2	555	6	US-10-029-386-5272, Ap
C 18	17.2	78.2	584	4	US-09-925-065A-475820, A
C 19	17.2	78.2	612	6	US-10-369-493-35681, A
C 20	17.2	78.2	754	3	US-09-910-943-130, Appl
C 21	17.2	78.2	1038	6	US-10-156-761-1058, Ap
C 22	17.2	78.2	1254	4	US-09-925-065A-56098, A
C 23	17.2	78.2	1254	4	US-09-925-065A-56099, A

C	24	17.2	78.2	1254	4	US-09-925-065A-56100	Sequence 56100, A
C	25	17.2	78.2	1348	7	US-10-437-963-36556	Sequence 36556, A
C	26	17.2	78.2	1867	7	US-10-302-172-749	Sequence 749, App
C	27	17.2	78.2	106664	5	US-10-175-523-97	Sequence 97, Appl
C	28	17.2	78.2	106664	10	US-11-099-266-97	Sequence 97, Appl
C	29	17.2	78.2	9025608	6	US-10-156-761-1	Sequence 1, Appl
C	30	16.8	76.4	156	7	US-10-437-963-16553	Sequence 16553, A
C	31	16.8	76.4	975	3	US-09-738-626-420	Sequence 420, App
C	32	16.8	76.4	1077	5	US-10-027-806-75	Sequence 75, Appl
C	33	16.8	76.4	1077	5	US-10-034-623-75	Sequence 75, Appl
C	34	16.8	76.4	1077	5	US-10-027-801-75	Sequence 75, Appl
C	35	16.8	76.4	1077	6	US-10-029-120-75	Sequence 75, Appl
C	36	16.8	76.4	1886	10	US-11-097-143-10247	Sequence 10247, A
C	37	16.8	76.4	2107	7	US-10-425-114-13350	Sequence 13350, A
C	38	16.8	76.4	2115	7	US-10-424-599-11032	Sequence 11032, A
C	39	16.8	76.4	4153	10	US-11-097-143-10246	Sequence 10246, A
C	40	16.8	76.4	32998	5	US-10-027-806-1	Sequence 1, Appl
C	41	16.8	76.4	32998	5	US-10-034-623-1	Sequence 1, Appl
C	42	16.8	76.4	32998	5	US-10-027-801-1	Sequence 1, Appl
C	43	16.8	76.4	32998	6	US-10-029-120-1	Sequence 1, Appl
C	44	16.8	76.4	238417	9	US-10-461-862-98	Sequence 98, Appl
C	45	16.8	76.4	3309400	3	US-09-738-626-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-925-065A-20834/c
; Sequence 20834, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20834
; LENGTH: 588
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-20834

Query Match 100.0%; Score 22; DB 4; Length 588;
Best Local Similarity 100.0%; Pred. No. 0.44; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;

QY 1 AGAAAGTGCCTGCGGTGCT 22
|||||
Db 67 AGAAAGTGCCTGCGGTGCT 46
|||||

RESULT 2
US-09-732-627A-1180/c
; Sequence 1180, Application US/09732627A
; Publication No. US2004012338A1
; GENERAL INFORMATION:
; APPLICANT: Fincher, Karen L.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants

```
; FILE REFERENCE: 38-21(51770)B
; CURRENT APPLICATION NUMBER: US/09/732,627A
; CURRENT FILING DATE: 2000-12-08
; NUMBER OF SEQ ID NOS: 4930
; SEQ ID NO 1180
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3493-057-P1-M1-D2
US-09-732-627A-1180

Query Match          79.1%; Score 17.4; DB 3; Length 381;
Best Local Similarity 94.7%; Pred. No. 98;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGAAGTGCCTGCGCGGT 19
    |||||
Db 37 AGAAGTGAACGCTGCGGT 19

RESULT 3
US-10-450-763-21563/c
; Sequence 21563, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 21563
; LENGTH: 1946
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIMILAR
; LOCATION: (944)..(1165)
; OTHER INFORMATION: 94% homologous to Escherichia coli trfE, accession number
; OTHER INFORMATION: U00039, Smith-Waterman Score=369.
US-10-450-763-21563

Query Match          79.1%; Score 17.4; DB 9; Length 1946;
Best Local Similarity 94.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AAGTATCGCTGCGCGGT 22
    |||||
Db 219 AAGTATCGCTGCGCGGT 201

RESULT 4
US-10-655-042-13
; Sequence 13, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13

Query Match          79.1%; Score 17.4; DB 7; Length 3307;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 AAGTATCGCTGCGCGGT 22
    |||||
Db 2833 AAGTATCGCTGCGCGGT 2851

RESULT 6
US-10-655-042-5
; Sequence 5, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 3334
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: E. coli RNase E deletion mutant
```

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (441)..(3296)
; OTHER INFORMATION:
US-10-655-042-5

Query Match      79.1%; Score 17.4; DB 7; Length 3334;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      4 AAGTGATCGCTGCCGTGGT 22
Db      2860 AAGTGATCGCTGCCGTGGT 2878

RESULT 7
US-10-655-042-7
; Sequence 7, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7
; LENGTH: 3388
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: E. coli RNase E deletion mutant
; NAME/KEY: CDS
; LOCATION: (441)..(3350)
; OTHER INFORMATION:
US-10-655-042-7

Query Match      79.1%; Score 17.4; DB 7; Length 3388;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      4 AAGTGATCGCTGCCGTGGT 22
Db      2914 AAGTGATCGCTGCCGTGGT 2932

RESULT 8
US-10-655-042-11
; Sequence 11, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11
; LENGTH: 3487
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: E. coli RNase E deletion mutant
; NAME/KEY: CDS
; LOCATION: (441)..(3449)
; OTHER INFORMATION:
US-10-655-042-11
```

```
Query Match      79.1%; Score 17.4; DB 7; Length 3487;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      4 AAGTGATCGCTGCCGTGGT 22
Db      3013 AAGTGATCGCTGCCGTGGT 3031

RESULT 9
US-10-655-042-3
; Sequence 3, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 3
; LENGTH: 3532
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: E. coli RNase E deletion mutant
; NAME/KEY: CDS
; LOCATION: (441)..(3494)
; OTHER INFORMATION:
US-10-655-042-3

Query Match      79.1%; Score 17.4; DB 7; Length 3532;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      4 AAGTGATCGCTGCCGTGGT 22
Db      3058 AAGTGATCGCTGCCGTGGT 3076

RESULT 10
US-10-655-042-9
; Sequence 9, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 9
; LENGTH: 3586
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: E. coli RNase E deletion mutant
; NAME/KEY: CDS
; LOCATION: (441)..(3548)
; OTHER INFORMATION:
US-10-655-042-9

Query Match      79.1%; Score 17.4; DB 7; Length 3586;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 4 AAGTGATCGCTGCCGTGGT 22
|||||
Db 3112 AAGTGATCGCTGCCGGGT 3130

RESULT 11
US-10-633-577-1
; Sequence 1, Application US/10633577
; Publication No. US20040029231A1
; GENERAL INFORMATION:
; APPLICANT: DREYFUS, MARC
; APPLICANT: LOPEZ, PASCAL
; TITLE OF INVENTION: RECOMBINANT POLYPEPTIDES
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING
; FILE REFERENCE: USB98APNREC
; CURRENT APPLICATION NUMBER: US/10/633,577
; CURRENT FILING DATE: 2003-08-05
; PRIOR APPLICATION NUMBER: PCT/FR99/01879
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: FR 98/10197
; PRIOR FILING DATE: 1998-08-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3661
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (441)..(3623)
US-10-633-577-1

Query Match 79.1%; Score 17.4; DB 7; Length 3661;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 AAGTGATCGCTGCCGTGGT 22
|||||
Db 3187 AAGTGATCGCTGCCGGGT 3205

RESULT 12
US-10-655-042-1
; Sequence 1, Application US/10655042
; Publication No. US20040126842A1
; GENERAL INFORMATION:
; APPLICANT: CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
; TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING RECOMBINANT
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: ECO
; CURRENT APPLICATION NUMBER: US/10/655,042
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3661
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (441)..(3623)
; OTHER INFORMATION:
US-10-655-042-1

Query Match 79.1%; Score 17.4; DB 7; Length 3661;
Best Local Similarity 94.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 4 AAGTGATCGCTGCCGTGGT 22
|||||
Db 3187 AAGTGATCGCTGCCGGGT 3205

RESULT 13
US-10-425-115-17122
; Sequence 17122, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 17122
; LENGTH: 359
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_115611C.1
US-10-425-115-17122

Query Match 78.2%; Score 17.2; DB 8; Length 359;
Best Local Similarity 86.4%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 AGAAAGTGATCGCTGCCGTGGT 22
|||||
Db 239 AGAAAGTGATCGCTGCCGTGAT 260

RESULT 14
US-09-864-761-21719/c
; Sequence 21719, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:02 ; Search time 4.556 Seconds
(without alignments)
8583.479 Million cell updates/sec

Title: US-09-555-529-18

Sequence: 1 agaaagtcgctgcctgcgtgcgt 22

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/1-COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5-COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A-COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B-COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H-COMB.seq:*
6: /cgn2_6/ptodata/1/ina/ECTUS-COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP-COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE-COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	23645	US-09-949-016-13916	Sequence 13916, A
2	17.4	79.1	3661	US-09-762-481B-1	Sequence 1, Appl1
3	17.2	78.2	1867	US-09-799-451-749	Sequence 749, App
4	16.8	76.4	968	US-09-605-703B-2171	Sequence 2171, Ap
5	16.8	76.4	1077	US-09-408-020-75	Sequence 75, Appl
6	16.8	76.4	32998	US-09-408-020-1	Sequence 1, Appl1
7	16.2	73.6	501	US-09-252-991A-13073	Sequence 13073, A
8	16.2	73.6	601	US-09-949-016-63710	Sequence 63710, A
9	16.2	73.6	605	US-08-630-915A-5	Sequence 5, Appl1
10	16.2	73.6	605	US-09-879-957-5	Sequence 5, Appl1
11	16.2	73.6	2586	US-09-107-532A-2231	Sequence 2331, Ap
12	16.2	73.6	2994	US-09-949-016-1858	Sequence 1858, Ap
13	16.2	73.6	15449	US-09-949-016-13600	Sequence 13600, A
14	16.2	73.6	42325	US-08-311-731A-111	Sequence 131, App
15	16.2	73.6	340380	US-09-949-016-14179	Sequence 14179, A
16	15.8	71.8	479	US-09-270-767-4961	Sequence 4961, Ap
17	15.8	71.8	479	US-09-270-767-20243	Sequence 20243, A
18	15.8	71.8	715	US-08-998-416-590	Sequence 590, App
19	15.8	71.8	4494	US-09-620-312D-861	Sequence 861, App
20	15.8	71.8	6263	US-08-781-802-3	Sequence 3, Appl1
21	15.8	71.8	26711	US-08-694-078-3	Sequence 3, Appl1
22	15.8	71.8	42931	US-09-949-016-14214	Sequence 14214, A
23	15.8	71.8	42931	US-08-311-731A-129	Sequence 129, App
24	15.8	71.8	4403765	US-09-103-840A-2	Sequence 2, Appl1

25	15.8	71.8	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
26	15.6	70.9	435	3	US-09-902-540-4149	Sequence 4149, Ap
27	15.6	70.9	437	3	US-09-270-767-11731	Sequence 11731, A
28	15.6	70.9	485	3	US-09-513-999C-13085	Sequence 13085, A
29	15.6	70.9	541	3	US-09-270-767-8408	Sequence 8408, Ap
30	15.6	70.9	541	3	US-09-270-767-23690	Sequence 23690, A
31	15.6	70.9	588	3	US-08-858-207A-250	Sequence 250, App
32	15.6	70.9	601	3	US-09-949-016-122813	Sequence 122813, A
33	15.6	70.9	601	3	US-09-949-016-130765	Sequence 130765, A
34	15.6	70.9	601	3	US-09-949-002-3739	Sequence 3739, Ap
35	15.6	70.9	601	3	US-09-949-002-9273	Sequence 9273, Ap
36	15.6	70.9	671	3	US-09-533-559-6307	Sequence 6307, Ap
37	15.6	70.9	732	3	US-09-252-991A-14582	Sequence 14582, A
38	15.6	70.9	1002	3	US-09-543-681A-447	Sequence 447, App
39	15.6	70.9	1356	3	US-09-270-767-10484	Sequence 10484, A
40	15.6	70.9	1401	3	US-09-489-039A-483	Sequence 483, App
41	15.6	70.9	1488	3	US-08-962-859A-1	Sequence 1, Appl1
42	15.6	70.9	1488	3	US-09-583-110-1905	Sequence 1905, Ap
43	15.6	70.9	1488	3	US-09-107-433-785	Sequence 785, App
44	15.6	70.9	1828	3	US-09-664-958-11	Sequence 11, Appl
45	15.6	70.9	1866	4	US-09-605-703B-603	Sequence 603, App

ALIGNMENTS

```

RESULT 1
US-09-949-016-13916/c
Sequence 13916, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13916
LENGTH: 23645
TYPE: DNA
ORGANISM: Human
US-09-949-016-13916
Query Match
Best Local Similarity 100.0%; Score 22; DB 3; Length 23645;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAAAGTATCGCTGCCGTGCT 22
DB 1789 AGAAAGTATCGCTGCCGTGCT 1768
RESULT 2
US-09-762-481B-1
Sequence 1, Application US/09762481B
Patent No. 6632639
GENERAL INFORMATION:
APPLICANT: LOPEZ, PASCAL
APPLICANT: DREYERUS, MARC
TITLE OF INVENTION: MUTANT E. COLI STRAINS, AND THEIR USE FOR PRODUCING
RECOMBINANT POLYPEPTIDES
FILE REFERENCE: US98APCNRCC
CURRENT APPLICATION NUMBER: US/09/762,481B
CURRENT FILING DATE: 2001-08-27
PRIOR APPLICATION NUMBER: PCT/FR99/01879

```

;; PRIOR FILING DATE: 1999-07-29
;; PRIOR APPLICATION NUMBER: FR 98/10197
;; PRIOR FILING DATE: 1998-08-07
;; NUMBER OF SEQ ID NOS: 6
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 3661
;; TYPE: DNA
;; ORGANISM: Escherichia coli
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (441)..(3623)
US-09-762-481B-1

Query Match 79.1%; Score 17.4; DB 3; Length 3661;
Best Local Similarity 94.7%; Pred. No. 64;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 AAGTATCGCTGCGGTGT 22
|||||
Db 3187 AAGTATCGCTGCGGTGT 3205

RESULT 3
US-09-799-451-749
;; Sequence 749, Application US/09799451
;; Patent No. 6783969
;; GENERAL INFORMATION:
;; APPLICANT: Tang, Y. Tom
;; APPLICANT: Zhou, Ping
;; APPLICANT: Goodrich, Ryle
;; APPLICANT: Asundi, Vinod
;; APPLICANT: Ren, Feiyan
;; APPLICANT: Zhang, Jie
;; APPLICANT: Xue, Aidong J.
;; APPLICANT: Zhao, Qing A.
;; APPLICANT: Wang, Jian-Rui
;; APPLICANT: Ma, Yundong
;; APPLICANT: Yamazaki, Victoria
;; APPLICANT: Chen, Rui-hong
;; APPLICANT: Wang, Zhilwei
;; APPLICANT: Wang, Dunfui
;; APPLICANT: Yang, Yonghong
;; APPLICANT: Wehrman, Tom
;; APPLICANT: Ghosh, Reena
;; APPLICANT: Drmanac, Radoje T.
;; TITLE OF INVENTION: No. 6783969e1 Nucleic Acids and
;; FILE REFERENCE: 803 Polypeptides
;; CURRENT APPLICATION NUMBER: US/09/799,451
;; CURRENT FILING DATE: 2001-03-05
;; NUMBER OF SEQ ID NOS: 948
;; SOFTWARE: pc_fl_genes Version 2.0
;; SEQ ID NO 749
;; LENGTH: 1867
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1103)..(1675)
US-09-799-451-749

Query Match 78.2%; Score 17.2; DB 3; Length 1867;
Best Local Similarity 86.4%; Pred. No. 73;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 AGAAGTATCGCTGCGGTGT 22
|||||
Db 1628 AGAAGTATCGCTGCGGTGT 1649

RESULT 4
US-09-605-703B-2171

;; Sequence 2171, Application US/09605703B
;; Patent No. 6962989
;; GENERAL INFORMATION:
;; APPLICANT: Pompejus, Markus
;; APPLICANT: Kroeger, Burkhard
;; APPLICANT: Schroder, Hartwig
;; APPLICANT: Zelder, Oskar
;; APPLICANT: Haberhauer, Gregor
;; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
;; FILE REFERENCE: BGI-129CP
;; CURRENT APPLICATION NUMBER: US/09/605,703B
;; CURRENT FILING DATE: 2000-06-27
;; PRIOR APPLICATION NUMBER: 60/142,764
;; PRIOR FILING DATE: 1999-07-08
;; PRIOR APPLICATION NUMBER: 60/152,318
;; PRIOR FILING DATE: 1999-09-03
;; NUMBER OF SEQ ID NOS: 2934
;; SEQ ID NO 2171
;; LENGTH: 968
;; TYPE: DNA
;; ORGANISM: Corynebacterium glutamicum
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(945)
;; OTHER INFORMATION: RXA01313
US-09-605-703B-2171

Query Match 76.4%; Score 16.8; DB 4; Length 968;
Best Local Similarity 90.0%; Pred. No. 11e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AGAAGTATCGCTGCGGTG 20
|||||
Db 665 AGAAGTATCGCTGCGGTG 684

RESULT 5
US-09-408-020-75/C
;; Sequence 75, Application US/09408020
;; Patent No. 6632937
;; GENERAL INFORMATION:
;; APPLICANT: Swanson, Ronald V.
;; APPLICANT: Feldman, Robert A.
;; APPLICANT: Schleper, Christa
;; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
;; FILE REFERENCE: DCRP.002A
;; CURRENT APPLICATION NUMBER: US/09/408,020
;; CURRENT FILING DATE: 1999-09-29
;; PRIOR APPLICATION NUMBER: 60/102,294
;; PRIOR FILING DATE: 1998-09-29
;; NUMBER OF SEQ ID NOS: 123
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 75
;; LENGTH: 1077
;; TYPE: DNA
;; ORGANISM: Cenarchaeum symbiosum
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(1077)
US-09-408-020-75

Query Match 76.4%; Score 16.8; DB 3; Length 1077;
Best Local Similarity 90.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 AAAGTATCGCTGCGGTGT 22
|||||
Db 828 AAAGTATCGCTGCGGTGT 809

RESULT 6
US-09-408-020-1

```

Sequence 1, Application US/09408020
Patent No. 6632937
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.
TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAETUM SYMBIOSUM
FILE REFERENCE: DCOIP.002A
CURRENT APPLICATION NUMBER: US/09/408,020
CURRENT FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: 60/102,294
PRIOR FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 123
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 32998
TYPE: DNA
ORGANISM: Cenarchaeum symbiosum
FEATURE:
NAME/KEY: CDS
LOCATION: (7604)...(8908)
FEATURE:
NAME/KEY: CDS
LOCATION: (8961)...(9767)
FEATURE:
NAME/KEY: CDS
LOCATION: (10545)...(10922)
FEATURE:
NAME/KEY: CDS
LOCATION: (13944)...(14612)
FEATURE:
NAME/KEY: CDS
LOCATION: (18638)...(20149)
FEATURE:
NAME/KEY: CDS
LOCATION: (20554)...(20955)
FEATURE:
NAME/KEY: CDS
LOCATION: (20956)...(21834)
FEATURE:
NAME/KEY: CDS
LOCATION: (25151)...(26377)
FEATURE:
NAME/KEY: CDS
LOCATION: (27535)...(28002)
FEATURE:
NAME/KEY: CDS
LOCATION: (28065)...(29483)
US-09-408-020-1

Query Match          76.4%; Score 16.8; DB 3; Length 32998;
Best Local Similarity 90.0%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 AAGTATCGCTGCGCGTGT 22
DB      26627 AAGTCTGCTGTCGCGTGT 26646

RESULT 7
US-09-252-991A-13073
Sequence 13073, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

```

```

PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 13073
LENGTH: 501
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13073

Query Match          73.6%; Score 16.2; DB 3; Length 501;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AGAAGTATCGCTGCGCGTGG 21
DB      251 AGAAGTATCATCTTCCGTGG 271

RESULT 8
US-09-949-016-63710/C
Sequence 63710, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 63710
LENGTH: 601
TYPE: DNA
ORGANISM: Human
US-09-949-016-63710

Query Match          73.6%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 AGAAGTATCGCTGCGCGTGG 21
DB      486 AGAGGCGATGCTGCGCGTGG 466

RESULT 9
US-08-630-915A-5/C
Sequence 5, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOWLER, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 605 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-630-915A-5

Query Match 73.6%; Score 16.2; DB 3; Length 605;
Best Local Similarity 85.7%; Pred. No. 2e+02; 3; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GAAAGTATGCTGCGCTGGT 22
Db 431 GCAAGTGTTCGCTGCCGGGT 411

RESULT 10
US-09-879-957-5/c
Sequence 5, Application US/09879957
Patent No. 6709821
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. 6709821h
KAY, Brian K.
FOWLES, Dana M.
MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 605 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-879-957-5

Query Match 73.6%; Score 16.2; DB 3; Length 605;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GAAAGTATGCTGCGCTGGT 22
Db 431 GCAAGTGTTCGCTGCCGGGT 411

RESULT 11
US-09-107-532A-2331/c
Sequence 2331, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 893-5007
TELEFAX: (781) 893-8277
INFORMATION FOR SEQ ID NO: 2331:
SEQUENCE CHARACTERISTICS:
LENGTH: 2586 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...2586
SEQUENCE DESCRIPTION: SEQ ID NO: 2331:
US-09-107-532A-2331

Query Match 73.6%; Score 16.2; DB 3; Length 2586;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAAGTATGCTGCTCCGTGT 22
DB 1122 GAAGTATGCTGCTCCGTGT 1102

RESULT 12
US-09-949-016-1858
Sequence 1858, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1858
LENGTH: 2994
TYPE: DNA
ORGANISM: Human
US-09-949-016-1858

Query Match 73.6%; Score 16.2; DB 3; Length 2994;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGAAGTATGCTGCTCCGTGT 21
DB 2317 AGAAGTATGCTGCTCCGTGT 2337

RESULT 13
US-09-949-016-13600
Sequence 13600, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 13600
LENGTH: 15449
TYPE: DNA
ORGANISM: Human
US-09-949-016-13600

Query Match 73.6%; Score 16.2; DB 3; Length 15449;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGAAGTATGCTGCTCCGTGT 21
DB 12771 AGAAGTATGCTGCTCCGTGT 12791

RESULT 14
US-08-311-731A-131/C
Sequence 131, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
DIAGNOSTICS AND THERAPEUTICS
TITLE OF INVENTION: 411
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSEE: WOLF, GREENFIELD & SACKS, P. C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 131:
SEQUENCE CHARACTERISTICS:
LENGTH: 42325 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIUM LAPRAE
US-08-311-731A-131

Query Match 73.6%; Score 16.2; DB 3; Length 42325;
Best Local Similarity 85.7%; Pred. No. 3.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AGAAGTATGCTGCTCCGTGT 21
DB 15586 AGGAGTATGCTGCTCCGTGT 15566

RESULT 15
US-09-949-016-14179
Sequence 14179, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755

```
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14179
; LENGTH: 340380
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) .. (340380)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14179
```

```
Query Match      73.6%; Score 16.2; DB 3; Length 340380;
Best Local Similarity 85.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      2  GAAAGTATCGCTGCCGTGGT 22
          |||||  |||||  |||||
Db      219584  GAAAGTGACAGCTGCCCTGGT 219604
```

```
Search completed: November 27, 2005, 01:10:47
Job time : 10.556 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:58:53 : Search time 122,977 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-18

Perfect score: 22

Sequence: 1 agaaatgacgcgtgcgtggt 22

Scoring table: IDENTITY_NUC

Gapop 10,0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_ges1.*
10: gb_ges2.*
11: gb_ges3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	22	100.0	529	5	BX504040 DKFZp686H
2	22	100.0	581	3	BP350798 BP350798
3	22	100.0	718	8	CX866441 HESCA 10
4	22	100.0	882	3	BQ221694 AGENCOURT
5	22	100.0	890	6	CD109094 AGENCOURT
6	22	100.0	922	6	CD389584 AGENCOURT
7	22	100.0	965	1	AL522388 AL522388
8	22	100.0	1339	4	CR618602 full-leng
9	21	95.5	552	7	CN256426 CN256426
10	20.4	92.7	951	7	C0725903 C0725903
11	20.4	92.7	988	7	C0726806 C0726806
12	20.4	92.7	1011	7	C0725532 C0725532
13	19.4	88.2	411	5	BY450707 BY450707
14	19	86.4	677	8	DR001039 DR001039
15	18	81.8	842	3	BI862461 BI862461
16	17.8	80.9	106	9	BH818205 BH818205
17	17.8	80.9	310	9	BH817324 BACPPI-G0
18	17.8	80.9	365	8	CV987860 lpcgrr1_1
19	17.8	80.9	370	5	BW888824 BW888824
20	17.8	80.9	453	7	CN211166 Rt fcd 04
21	17.8	80.9	508	9	BH825901 BACPPI-H
22	17.8	80.9	529	5	BU697159 LL21n1204

23	17.8	80.9	530	10	CZ633455	CZ633455 OM_Ba017
24	17.8	80.9	554	9	BH828298	BH828298 BACPP27-D
25	17.8	80.9	591	10	CZ611198	CZ611198 OM_Ba013
26	17.8	80.9	633	8	CV993795	CV993795 IPCGF+3.1
27	17.8	80.9	649	6	CB939232	CB939232 IPCGX13
28	17.8	80.9	650	8	CV996304	CV996304 IPCGF+4.1
29	17.8	80.9	660	6	CB937334	CB937334 IPCGX13
30	17.8	80.9	668	7	CF947745	CF947745 UI-D-QC0-
31	17.8	80.9	694	8	CV994840	CV994840 IPCGF+3.1
32	17.8	80.9	713	5	BU221080	BU221080 603755374
33	17.8	80.9	713	9	BH835596	BH835596 BACPP8-H1
34	17.8	80.9	754	10	CM673179	CM673179 OB_Ba006
35	17.8	80.9	763	3	BJ050799	BJ050799 BJ050799
36	17.8	80.9	776	10	CZ675529	CZ675529 OM_Ba023
37	17.8	80.9	789	5	BU450710	BU450710 603217766
38	17.8	80.9	896	10	CNS02CM4	AL191677 Tetraodon
39	17.8	80.9	925	7	CK410468	CK410468 AUF_IpHdk
40	17.4	79.1	115	6	CF557778	CF557778 1115031C0
41	17.4	79.1	116	6	CF561674	CF561674 1115067D1
42	17.4	79.1	163	7	CV406745	CV406745 RCO-TN008
43	17.4	79.1	395	1	AU050627	AU050627 AU050627
44	17.4	79.1	432	5	BY461925	BY461925 BY461925
45	17.4	79.1	543	10	CL978385	CL978385 OsIFCC031

ALIGNMENTS

RESULT 1
BX504040
LOCUS
DEFINITION BX504040 529 bp mRNA linear EST 04-SEP-2003
DKFZp686H04126 r1 686 (synonym: hlcc3) Homo sapiens CDNA clone
DKFZp686H04126 5', mRNA sequence.
ACCESSION BX504040
VERSION BX504040.1 GI:32028491
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 529)
Poustka.A., Albert.R., Moosmayer.P., Schupp.I., Wellenreuther.R.,
Mewes.H.W., Weil.B., Amid.C., Osanger.A., Fobo.G., Han.M. and
Wiemann.S.
EST (Poustka.A., Albert.R., Moosmayer.P., Schupp.I., Wellenreuther.R.,
Wellenreuther.R., et al.)
JOURNAL Unpublished (2003)
COMMENT Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Sequenced by DKFZ (German Cancer Research Center,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No sl sequence available.
This clone (DKFZp686H04126) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
Location/Qualifiers
1..529
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686H04126"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="vector: pTriplex2; Site_1: sfIIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

```

Query Match      100.0%; Score 22; DB 5; Length 529;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGTGCCTGCGCGTGGT 22
    |||||
Db 36 AGAAGTGCCTGCGCGTGGT 57

RESULT 2
LOCUS BP350798
DEFINITION BP350798 Sugano cDNA library, squamous cell TE13 Homo sapiens cDNA
clone T3R01993, mRNA sequence.
ACCESSION BP350798
VERSION BP350798.1 GI:52280784
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 581)
AUTHORS Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,
Mizushima-Sugano,J., Nakai,K. and Sugano,S.
TITLE Sequence comparison of human and mouse genes reveals a homologous
block structure in the promoter regions
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)
PUBMED 15342556
COMMENT Contact: Yutaka Suzuki
Department of Virology
Institute of Medical Science, University of Tokyo
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
Email: yusuzuki@ims.u-tokyo.ac.jp.
Location/Qualifiers
1..581
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="T3R01993"
/cell_type="squamous cell"
/cell_line="TE13"
/clone_lib="Sugano cDNA library, squamous cell TE13"
/notes="well-differentiated squamous cell carcinoma"

FEATURES
source
1..581
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:7472092"
/sex="male"
/tissue_type="embryonic stem"
/cell_type="human embryonic stem cells"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="NIH MGC 262"
/notes="Vector: pExpress-1; Site 1: NotI; Site 2: EcoRV;
RNA obtained from human embryonic stem cells isolated from
the inner cell mass of blastocyst stage embryos and
differentiated to an early neural progenitor cell type.
Cell line id and NIH Registry designation is BG01.
Positive for Nestin and Musashi expression. Passage number
18. CDNA primed using oligo-dT primer:
5'-GCATTAGTCTAGATCGGAGCGCGGCC(T)25-3' and cloned into
the EcoRV/NotI sites of pExpress-1. This primary library
is non-normalized (normalized primary library is
NIH MGC 259). It was constructed by Express Genomics
(Frederick, MD). Sequence ends have been trimmed to
exclude vector and regions below phred quality 16. Note:
this is a Mammalian Gene Collection library."

ORIGIN
Query Match      100.0%; Score 22; DB 8; Length 718;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGTGCCTGCGCGTGGT 22
    |||||
Db 2 AGAAGTGCCTGCGCGTGGT 23

RESULT 4
LOCUS BQ221694
DEFINITION AGENCOURT_7549812 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6057750
5', mRNA sequence.
ACCESSION BQ221694
VERSION BQ221694.1 GI:20403094
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 882)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov

```

Other_ESTs: HES4_10.H01.bi.A037
Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs@mail.nih.gov
Tissue Procurement: BresaGen, Inc.
cDNA Library Preparation: Express Genomics, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Laboratory for Genomics and Bioinformatics,
University of Georgia
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM15776 row: o column: 02
Seq primer: JENREV (CAGGAACAGCATGACC)
High quality sequence stop: 718.

FEATURES

source
1..718
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:7472092"
/sex="male"
/tissue_type="embryonic stem"
/cell_type="human embryonic stem cells"
/lab_host="DH10B-T1 phage-resistant E. coli"
/clone_lib="NIH MGC 262"
/notes="Vector: pExpress-1; Site 1: NotI; Site 2: EcoRV;
RNA obtained from human embryonic stem cells isolated from
the inner cell mass of blastocyst stage embryos and
differentiated to an early neural progenitor cell type.
Cell line id and NIH Registry designation is BG01.
Positive for Nestin and Musashi expression. Passage number
18. CDNA primed using oligo-dT primer:
5'-GCATTAGTCTAGATCGGAGCGCGGCC(T)25-3' and cloned into
the EcoRV/NotI sites of pExpress-1. This primary library
is non-normalized (normalized primary library is
NIH MGC 259). It was constructed by Express Genomics
(Frederick, MD). Sequence ends have been trimmed to
exclude vector and regions below phred quality 16. Note:
this is a Mammalian Gene Collection library."

ORIGIN

Query Match 100.0%; Score 22; DB 8; Length 718;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGAAGTGCCTGCGCGTGGT 22
 |||||
Db 2 AGAAGTGCCTGCGCGTGGT 23

RESULT 4
LOCUS BQ221694
DEFINITION AGENCOURT_7549812 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6057750
5', mRNA sequence.
ACCESSION BQ221694
VERSION BQ221694.1 GI:20403094
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 882)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov

Tissue Procurement: DCTD/DTP/Gazdar
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LHAM1321 row: 1 column: 07
 High quality sequence stop: 581.
 Location/Qualifiers
 1. :882
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6057750"
 /tissue_type="large cell carcinoma"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC 68"
 /note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.8 kb. Library constructed by Life Technologies."

ORIGIN

Query Match 100.0%; Score 22; DB 3; Length 882;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGTGCCTGCGCGTGGT 22
 |||||
 Db 5 AGAAGTGCCTGCGCGTGGT 26

RESULT 5

CD109094
 LOCUS
 DEFINITION AGENCOURT 14019646 NIH_MGC_179 Homo sapiens cDNA clone
 IMAGE:30366888 5', mRNA sequence.
 CD109094
 VERSION
 EST.

QY 1 AGAAGTGCCTGCGCGTGGT 22
 |||||
 Db 5 AGAAGTGCCTGCGCGTGGT 26

ORIGIN

Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Homo.

REFERENCE 1 (bases 1 to 890)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein
 cDNA Library Preparation: Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: NDAM437 row: d column: 01
 High quality sequence stop: 653.
 Location/Qualifiers
 1. :890

FEATURES

source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30366888"
 /tissue_type="Pituitary"
 /lab_host="DH10B-Ton A (T1 and T5 phage resistances)"
 /clone_lib="NIH_MGC 179"
 /note="Organ: brain; Vector: pCMV-SPORT6.1; Site 1: EcoRV (destroyed); Site 2: NotI; Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon

cloning). Average insert size 1.1 kb. Library was constructed by (Invitrogen). Note: this is a NIH_MGC Library."

ORIGIN
 Query Match 100.0%; Score 22; DB 6; Length 890;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGTGCCTGCGCGTGGT 22
 |||||
 Db 22 AGAAGTGCCTGCGCGTGGT 43

RESULT 6

CD389584
 LOCUS
 DEFINITION AGENCOURT 14304876 NIH_MGC_173 Homo sapiens cDNA 5', mRNA sequence.
 CD389584
 ACCESSION
 VERSION
 EST.

QY 1 AGAAGTGCCTGCGCGTGGT 22
 |||||
 Db 22 AGAAGTGCCTGCGCGTGGT 43

ORIGIN

Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homiidae; Homo.

REFERENCE 1 (bases 1 to 922)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jamie Thompson, University of WI
 cDNA Library Preparation: Gina Zastrow-Hayes
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: NDKM67 row: n column: 11
 High quality sequence start: 28
 High quality sequence stop: 601.

Location/Qualifiers
 1. :922
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="embryonic trophoblasts, made from WA01 stem cells"

Plate: NDKM67 row: n column: 11
 High quality sequence start: 28
 High quality sequence stop: 601.

FEATURES

source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="embryonic trophoblasts, made from WA01 stem cells"
 /lab_host="DH10B Tona"
 /clone_lib="NIH_MGC 173"
 /note="Vector: pDONR201; Site 1: attP2; Site 2: attP1;
 LIBR PRIMING - oligo dT; METHOD - full-length enriched;
 LIBR_PROVIDER - Bradfield"

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 922;
 Best Local Similarity 100.0%; Pred. No. 7;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGTGCCTGCGCGTGGT 22
 |||||
 Db 38 AGAAGTGCCTGCGCGTGGT 59

RESULT 7

AL522388
 LOCUS
 DEFINITION AL522388 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens cDNA clone CS0DB008YGI8 5-PRIME, mRNA sequence.

```

ACCESSION   AL522388
VERSION     AL522388.3  GI:45697719
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 965)
AUTHORS    Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished (2001)
COMMENT    On Feb 13, 2001 this sequence version replaced gi:31040660.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
5543.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DB008B09QP1&c=5543.r.
FEATURES             source
    Location/Qualifiers
        1..965
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DB008YG18"
            /tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"
            /notes="1st strand cDNA was primed with a NotI-oligo(dT)
            primer. Five prime end enriched, double-strand cDNA was
            digested with Not I and EcoR V sites of the pCMVSPORT 6
            vector. Library was normalized."
ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 965;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAAGTGCCTGCGCGTGGT 22
    |||||
Db 41 AGAAAGTGCCTGCGCGTGGT 62

RESULT 8
CR618602
LOCUS       full-length cDNA clone CS0DB008YG18 of Neuroblastoma Cot
DEFINITION  10-normalized of Homo sapiens (human).
ACCESSION   CR618602
VERSION     CR618602.1  GI:504999409
KEYWORDS    HTC; CNSLT_CDNA.
SOURCE      Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 1339)
AUTHORS    Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished
REMARK     Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue
            2 (bases 1 to 1339)
            Genoscope.
            Direct Submission
            Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

ACCESSION   AL522388
VERSION     AL522388.3  GI:45697719
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 965)
AUTHORS    Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished (2001)
COMMENT    On Feb 13, 2001 this sequence version replaced gi:31040660.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
5543.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DB008B09QP1&c=5543.r.
FEATURES             source
    Location/Qualifiers
        1..965
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DB008YG18"
            /tissue_type="NEUROBLASTOMA COT 10-NORMALIZED"
            /notes="1st strand cDNA was primed with a NotI-oligo(dT)
            primer. Five prime end enriched, double-strand cDNA was
            digested with Not I and EcoR V sites of the pCMVSPORT 6
            vector. Library was normalized."
ORIGIN
Query Match      100.0%; Score 22; DB 1; Length 965;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAAGTGCCTGCGCGTGGT 22
    |||||
Db 41 AGAAAGTGCCTGCGCGTGGT 62

RESULT 8
CR618602
LOCUS       full-length cDNA clone CS0DB008YG18 of Neuroblastoma Cot
DEFINITION  10-normalized of Homo sapiens (human).
ACCESSION   CR618602
VERSION     CR618602.1  GI:504999409
KEYWORDS    HTC; CNSLT_CDNA.
SOURCE      Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 1339)
AUTHORS    Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished
REMARK     Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue
            2 (bases 1 to 1339)
            Genoscope.
            Direct Submission
            Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES             source
    Location/Qualifiers
        1..1339
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DB008YG18"
            /tissue_type="Neuroblastoma Cot 10-normalized"
            /plasmid="pCMVSPORT_6"
ORIGIN
Query Match      100.0%; Score 22; DB 4; Length 1339;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAAGTGCCTGCGCGTGGT 22
    |||||
Db 41 AGAAAGTGCCTGCGCGTGGT 62

RESULT 9
CN256426
LOCUS       17000600085458 GRN_PREHEP Homo sapiens cDNA 5', mRNA sequence.
DEFINITION  552 bp mRNA linear EST 16-MAY-2004
ACCESSION   CN256426
VERSION     CN256426.1  GI:47272840
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 552)
AUTHORS    Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Flsek, G.J.,
            Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R.,
            Lebkowski, J. and Stanton, L.W.
TITLE      Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
JOURNAL    Nat. Biotechnol. 22 (6), 707-716 (2004)
PUBMED     15146197
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 552 Std Error: 0.00.
FEATURES             source
    Location/Qualifiers
        1..552
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /tissue_type="embryonic stem cells, DMSO-treated H9 cell
            line"
            /clone_lib="GRN PREHEP"
            /note="oligo dT primed, full-length enriched cDNA library
            from DMSO-treated HES cell line H9 (p22) maintained in
            feeder-free conditions"
ORIGIN
Query Match      95.5%; Score 21; DB 7; Length 552;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GAAAGTGCCTGCGCGTGGT 22
    |||||
Db 9 GAAAGTGCCTGCGCGTGGT 29

```


REFERENCE AUTHORS	1 (bases 1 to 1011) Magness,C.L., Fellin,P.C., Thomas,M.J., Korth,M.J., Agy,M.B., Prolli,S.C., Fitzgibbon,M., Scherer,C.A., Miner,D.G., Katze,M.G. and Iadonato,S.P.	
TITLE	Analysis of the Macaca mulatta transcriptome and the sequence divergence between Macaca and human	
JOURNAL PUBMED	Genome Biol. 6 (7), R60 (2005) 15998449	
COMMENT	Contact: C. Magness Illumigen Biosciences Inc. 2203 Airport Way S, Suite 450, Seattle, WA 98134, USA Tel: 2063780400 Fax: 2063780408 Email: cmagness@illumigen.com Sequenced on 2004.05.11. 622 Q20 bases. Library Preparation: Prof. Michael Katze Lab at University of Washington DNA Sequencing: Illumigen Biosciences Inc. For further information, see http://www.macaque.org PCR Primers FORWARD: CCTCCTACTAAAGGGAACAAA BACKWARD: CACTATAGGCGAATTGGGTA Insert Length: 1011 Std Error: 0.00 Plate: CL000250 row: E column: 09 Seq primer: CCTCCTACTAAAGGGAACAAA POLYA=Yes.	
FEATURES source	Location/Qualifiers	
	1..1011	
	/organism="Macaca mulatta"	
	/mol_type="mRNA"	
	/strain="Indian"	
	/db_xref="taxon:9544"	
	/clone="IBIUIW:28218"	
	/sex="female"	
	/dev_stage="adult"	
	/lab_host="E. coli SOLR"	
	/clone_lib="Katze WMBR"	
	/note="Organ: brain; Vector: Uni-ZAP XR; Site 1: EcoR I; Site 2: Xho I; Created from StrataGene ZAP-CDNA Synthesis kit (Catalog #200400) and ZAP-CDNA Gigapack III Gold Cloning Kit (Catalog #200450)"	
ORIGIN	Query Match 92.7%; Score 20.4; DB 7; Length 1011; Best Local Similarity 95.5%; Pred. No. 45; Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 AGAAAGTGATCGTGGCGTGGT 22	
Db	25 AGAAAGTGACCGTGGCGTGGT 46	
RESULT 13	BY450707	411 bp mRNA linear EST 13-DEC-2002
LOCUS	BY450707	musculus cDNA clone K630047B05 3', mRNA sequence.
DEFINITION	BY450707	
ACCESSION	BY450707.1	GI:26744401
VERSION		
KEYWORDS		
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
REFERENCE AUTHORS	Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Kiyosawa,H., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustinich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Sempile,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verdaro,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 12466851 Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/ Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site (http://genome.gsc.riken.go.jp) for further details.	
FEATURES source	Location/Qualifiers	
	1..411	
	/organism="Mus musculus"	
	/mol_type="mRNA"	
	/strain="C57BL/6J"	
	/db_xref="taxon:10090"	
	/clone="K630047B05"	
	/tissue_type="whole body"	
	/dev_stage="14 days embryo"	
	/clone_lib="RIKEN full-length enriched, 14 days embryo whole body"	
ORIGIN	Query Match 88.2%; Score 19.4; DB 5; Length 411; Best Local Similarity 95.2%; Pred. No. 1.2e+02; Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 AGAAAGTGATCGTGGCGTGG 21	

```

Db      31 AGAAGTGATAGCTGCGGTGG 51
|||||
RESULT 14
LOCUS   DR001039
DEFINITION
TC112658 Human fetal brain, large insert, pCMV expression library
Homo sapiens cDNA clone TC112658 5' similar to Homo sapiens KIN,
antigenic determinant of recA protein homolog (mouse) (KIN), mRNA
sequence.
ACCESSION
DR001039
VERSION  DR001039.1  GI:66260912
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE
1 (bases 1 to 677)
Birkett,C., Cho,J., Gau,Y., Hamer,R., Kelly,S., Kovacs,K., Liu,L.,
Liu,X., Porter,J., Sachs,A., Shu,Y., Sun,Z., Wong,J., Wu,M.,
Zhang,X., Jay,G. and He,W.
High-throughput cloning of full-length human cDNAs directly from
cDNA libraries optimized for large and rare transcripts
Unpublished (2005)
JOURNAL
COMMENT
Contact: Kovacs, KF
High Throughput cDNA Cloning
Origene Technologies, Inc. ( www.origene.com )
6 Taft Court, Suite 100, Rockville, MD 20850, USA
Tel: 301 340 3188
Fax: 301 340 8606
Email: cDNA@origene.com
This EST submission is part of an on-going human full-length
cloning project at Origene Technologies, Inc.
Please contact Origene for access.
Origene Technologies, Inc.
6 Taft Ct. Suite 100
Rockville, MD 20850
Tel: (301) 340-3188
http://www.origene.com
Seg primer: pCMV6 5prime forward vector primer, Origene
Technologies Inc.
FEATURES
source
1..677
/organism="Homo sapiens"
/mol_type="mRNA"
/db|xref="taxon:9606"
/clone="TC112658"
/tissue_type="Fetal Brain"
/clone_lib="Human fetal brain, large insert, pCMV
expression library"
/notes="Organ: Fetal Brain; Vector: pCMV6-XL4; Site 1:
EcoRI; Site 2: XhoI/SalI compatible end ligatio; Oligo-dT
primed reverse transcription optimized for large and GC
rich mRNA transcripts, cDNA size selection, optimized
ligation for large inserts into mammalian expression
vector, random clones selected for end sequence
verification of full-length genes"
ORIGIN
Query Match      86.4%; Score 19; DB 8; Length 677;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      4 AAGTGATCGTGCCTGGT 22
|||||
Db      26 AAGTGATCGTGCCTGGT 44
|||||

RESULT 15
LOCUS   BI862461
DEFINITION
603391378F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5406499 5',
mRNA sequence.
ACCESSION
BI862461
VERSION  BI862461.1  GI:16003208
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE
1 (bases 1 to 842)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue procurement: DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12036 row: 1 column: 20
High quality sequence stop: 834.
FEATURES
source
1..842
/organism="Homo sapiens"
/mol_type="mRNA"
/db|xref="taxon:9606"
/clone="IMAGE:5406499"
/tissue_type="mammary adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_87"
/notes="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
ORIGIN
Query Match      81.8%; Score 18; DB 3; Length 842;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AGAAGTGATCGTGCCTGGC 18
|||||
Db      23 AGAAGTGATCGTGCCTGGC 40
|||||

Search completed: November 27, 2005, 00:57:53
Job time : 125.977 secs

```

This Page Blank (usp1c)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 14.1112 Seconds
(without alignments)
10390.557 Million cell updates/sec

Title: US-09-555-529-18
Perfect score: 22
Sequence: 1 agaaagtgcgtgcgtggt 22

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002as.*
7: Geneseq2002bs.*
8: Geneseq2003as.*
9: Geneseq2003bs.*
10: Geneseq2003cs.*
11: Geneseq2003ds.*
12: Geneseq2004as.*
13: Geneseq2004bs.*
14: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	2	AAx85566 PCR prime
2	22	100.0	1002	2	AAx85551 CDNA of a
3	22	100.0	1296	2	AAx85549 CDNA of a
C 4	17.4	79.1	381	12	ADP92169 Cotton ex
C 5	17.4	79.1	1946	5	AAx85759 DNA encod
6	17.4	79.1	3661	3	AAx291119 Mutated E
7	17.4	79.1	3661	3	AAx291118 Escherich
C 8	17.2	78.2	412	4	AAx24336 Probe #14
C 9	17.2	78.2	412	4	ABA69474 Human fo
C 10	17.2	78.2	412	4	AAx49606 Probe #18
C 11	17.2	78.2	412	4	ABA51446 Human bre
C 12	17.2	78.2	412	4	ABA36399 Probe #14
C 13	17.2	78.2	412	4	AAx43550 Human bon
C 14	17.2	78.2	412	4	AAx17729 Human bra
C 15	17.2	78.2	412	4	ABA43195 Human liv
C 16	17.2	78.2	412	5	AAx09878 Probe #98
C 17	17.2	78.2	412	6	ABA17689 Human gen
C 18	17.2	78.2	421	4	AAx15156 Probe #50
C 19	17.2	78.2	421	4	ABA56909 Human fo

C 20	17.2	78.2	421	4	AAx136484
C 21	17.2	78.2	421	4	ABA46344 Human bre
C 22	17.2	78.2	421	4	ABA26524 Probe #49
C 23	17.2	78.2	421	4	AAx30540 Human bon
C 24	17.2	78.2	421	4	AAx05006 Human bra
C 25	17.2	78.2	421	4	ABA30203 Human liv
C 26	17.2	78.2	421	5	AAx04899 Probe #48
C 27	17.2	78.2	421	6	ABA05185 Human gen
C 28	17.2	78.2	421	6	AAx99223 Human pro
C 29	17.2	78.2	475	14	AAx20832 Novel hum
C 30	17.2	78.2	555	12	ACH72077 Human gen
C 31	17.2	78.2	612	13	ADx60007 Bacterial
C 32	17.2	78.2	754	6	ABx76882 Frog embr
C 33	17.2	78.2	1867	6	ABx211867 Human pol
C 34	17.2	78.2	1867	12	ADx44385 Novel hum
C 35	17.2	78.2	106664	13	ADx35021 Murine cd
C 36	16.8	76.4	457	4	AAx56789 Human imm
C 37	16.8	76.4	975	5	AAx65385 C. glutami
C 38	16.8	76.4	975	8	ACA02093 C. glutam
C 39	16.8	76.4	1077	3	AAx55224 C. symbio
C 40	16.8	76.4	1081	4	AAx68843 Human imm
C 41	16.8	76.4	1081	4	AAx68842 Human imm
C 42	16.8	76.4	1886	4	ABL08671 Drosophil
C 43	16.8	76.4	2107	13	ADx30530 Plant ful
C 44	16.8	76.4	4153	4	ABL08670 Drosophil
C 45	16.8	76.4	32998	3	AAx55186 Cenarchae

ALIGNMENTS

RESULT 1
AAx85566

ID AAx85566 standard; cDNA; 22 BP.

XX AAx85566;

DT 07-SEP-1999 (first entry)

DE PCR primer and probe used to detect human kin17 nucleic acids.

KW Human; kin17 protein; cell proliferation; fertility; probe;

KW hyperproliferative disease; protein interaction; curved DNA;

KW HIV replication; HIV integration; repair enzyme; PCR primer; ss.

XX Synthetic.

OS Homo sapiens.

PN FR2772046-A1.

XX PD 11-JUN-1999.

XX PF 09-DEC-1997; 97FR-00015536.

XX PR 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell

XX proliferation or fertility.

XX Claim 7; Page 5; 69pp; French.

XX AAx5553-69 represent PCR primers and probes used to detect nucleic acids
encoding human kin17 protein. The mammalian kin17 protein is useful for
preparing a medicament for controlling cell proliferation or for
controlling fertility. The medicaments can also be used to treat
hyperproliferative diseases. Fragments between amino acids 55 and 235
(preferably between amino acids 129 and 228) of a mammalian kin17 protein

CC are useful for regulating the interaction between proteins and curved
 CC DNA. The fragment can be used to block replication of HIV or its
 CC integration into the human genome or to target repair enzymes to curved
 CC DNA sites. Expression vectors for kin17 can be used for controlling cell
 CC proliferation

SQ Sequence 22 BP; 5 A; 4 C; 8 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 2; Length 22;
 Best Local Similarity 100.0%; Pred. No. 0.49;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAAGTGCCTGCGTGGT 22
 |||||
 Db 1 AGAAGTGCCTGCGTGGT 22

RESULT 2

AAx85551
 ID AAX85551 standard; cDNA; 1002 BP.

XX AC AAX85551;

XX DT 07-SEP-1999 (first entry)

XX DE cDNA of a gene coding for a truncated human kin17 protein.

XX KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme; ss.

XX OS Homo sapiens.

XX PN FR2772046-A1.

XX PD 11-JUN-1999.

XX PF 09-DEC-1997; 97FR-00015536.

XX PR 09-DEC-1997; 97FR-00015536.

XX PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX DR WPI; 1999-359999/31.

XX PT New DNA coding for human kin17 protein - useful for controlling cell
 PT proliferation or fertility.

XX PS Claim 5; Page 32; 69pp; French.

XX CC The present sequence encodes a truncated human kin17 protein with amino
 CC acids 129-228 deleted. The mammalian kin17 protein is useful for
 CC preparing a medicament for controlling cell proliferation or for
 CC controlling fertility. The medicaments can also be used to treat
 CC hyperproliferative diseases. Fragments between amino acids 55 and 235
 CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein
 CC are useful for regulating the interaction between proteins and curved
 CC DNA. The fragment can be used to block replication of HIV or its
 CC integration into the human genome or to target repair enzymes to curved
 CC DNA sites. Expression vectors for kin17 can be used for controlling cell
 CC proliferation

XX SQ Sequence 1002 BP; 343 A; 189 C; 230 G; 240 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 2; Length 1002;
 Best Local Similarity 100.0%; Pred. No. 0.81;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAAGTGCCTGCGTGGT 22
 |||||
 Db 32 AGAAGTGCCTGCGTGGT 53

RESULT 3
 AAX85549
 ID AAX85549 standard; cDNA; 1296 BP.

XX AC AAX85549;

XX DT 07-SEP-1999 (first entry)

XX DE cDNA of a gene coding for the human kin17 protein.

XX KW Human; kin17 protein; cell proliferation; fertility;
 KW hyperproliferative disease; protein interaction; curved DNA;
 KW HIV replication; HIV integration; repair enzyme; ss.

XX OS Homo sapiens.

XX PN FR2772046-A1.

XX PD 11-JUN-1999.

XX PF 09-DEC-1997; 97FR-00015536.

XX PR 09-DEC-1997; 97FR-00015536.

XX PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX DR WPI; 1999-359999/31.

XX PT New DNA coding for human kin17 protein - useful for controlling cell
 PT proliferation or fertility.

XX PS Claim 1; Page 30-31; 69pp; French.

XX CC The present sequence encodes a human kin17 protein. The mammalian kin17
 CC protein is useful for preparing a medicament for controlling cell
 CC proliferation or for controlling fertility. The medicaments can also be
 CC used to treat hyperproliferative diseases. Fragments between amino acids
 CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
 CC kin17 protein are useful for regulating the interaction between proteins
 CC and curved DNA. The fragment can be used to block replication of HIV or
 CC its integration into the human genome or to target repair enzymes to
 CC curved DNA sites. Expression vectors for kin17 can be used for
 CC controlling cell proliferation

XX SQ Sequence 1296 BP; 453 A; 240 C; 310 G; 293 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 2; Length 1296;
 Best Local Similarity 100.0%; Pred. No. 0.84;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAAGTGCCTGCGTGGT 22
 |||||
 Db 32 AGAAGTGCCTGCGTGGT 53

RESULT 4

ADP92169/c
 ID ADP92169 standard; cDNA; 381 BP.

XX AC ADP92169;

XX DT 09-SEP-2004 (first entry)

XX DE Cotton expressed sequence tag, EST, #1180.

XX KW Cotton; ss; EST; expressed sequence tag; plant; plant protection;
 KW plant improvement; marker-assisted breeding.

XX OS Gossypium hirsutum; variety Nucleon33B.

XX PN US2004123338-A1.
 XX PD 24-JUN-2004.
 XX PP 08-DEC-2000; 2000US-00732627.
 XX PR 10-DEC-1999; 99US-0170255P.
 XX PA (FINC/) FINCHER K L.
 XX PI Fincher KL;
 XX P1 WPI; 2004-479807/45.
 XX DR
 XX PT New substantially purified nucleic acid molecule that encodes a cotton
 PT protein or its fragment, useful as molecular tool for the targeting and
 PT isolation of novel genes for plant protection and improvement.
 XX
 XX PS Claim 1; SEQ ID NO 1180; 30pp; English.
 XX
 CC The invention relates to a substantially purified nucleic acid molecule
 CC that encodes a cotton protein or its fragment comprising an EST
 CC (expressed sequence tag) appearing as ADP9090-ADP95919. Also included
 CC are a substantially purified cotton protein or its fragment encoded by a
 CC nucleic acid molecule above and a transformed plant (having a nucleic
 CC acid molecule which comprises: an exogenous promoter region which
 CC functions in a plant cell to cause the production of a mRNA molecule; a
 CC structural nucleic acid molecule comprising one of the ESTs or their
 CC complements; a 3' non-translated sequence that functions in the plant
 CC cell to cause termination of transcription and addition of polyadenylated
 CC ribonucleotides to a 3' end of the mRNA molecule). The ESTs are useful as
 CC molecular tool for the targeting and isolation of novel genes for plant
 CC protection and improvement. The ESTs are useful for developing new
 CC strategies for understanding critical plant developmental and metabolic
 CC pathways, for isolating genes and promoters, for identifying and mapping
 CC the genes involved in developmental and metabolic pathways, and for
 CC determining gene function. The cotton nucleic acid molecules are useful
 CC as molecular tags to isolate genetic regions, isolate genes, map genes,
 CC and determine gene function. The nucleic acid molecules are useful for
 CC determining if genes are members of a particular gene family and for use
 CC in marker-assisted breeding programs. The present sequence is one of the
 CC 4930 cotton ESTs of the invention. NOTE: The sequences are not displayed
 CC in the specification but are available in electronic format from the
 CC USPTO at seqdata.uspto.gov/sequence.html?docID=20040123338.
 XX
 XX SQ Sequence 381 BP; 99 A; 85 C; 114 G; 83 T; 0 U; 0 Other;
 SQ
 Query Match 79.1%; Score 17.4; DB 12; Length 381;
 Best Local Similarity 94.7%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AGAAGTGTATCGCTGCCGT 19
 Db 37 AGAAGTGTATCGCTGCCGT 19
 RESULT 5
 ID AAS85759/c
 XX ID AAS85759 standard; cDNA; 1946 BP.
 XX AC AAS85759;
 XX DT 13-FEB-2002 (first entry)
 XX DE DNA encoding novel human diagnostic protein #21563.
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX OS Homo sapiens.
 XX FH WO200175067-A2.
 XX FT

XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US008631.
 XX PR 31-MAR-2000; 2000US-00540217.
 XX PR 23-AUG-2000; 2000US-00649167.
 XX PA (HYSB-) HYSB INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX P1 WPI; 2001-639362/73.
 XX DR P-PSDB; ABG21572.
 XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 XX PS Claim 1; SEQ ID NO 21563; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (I) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX SQ Sequence 1946 BP; 423 A; 516 C; 519 G; 488 T; 0 U; 0 Other;
 SQ
 Query Match 79.1%; Score 17.4; DB 5; Length 1946;
 Best Local Similarity 94.7%; Pred. No. 1.8e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 4 AAGTGATCGCTGCCGTGCT 22
 Db 219 AAGTGATCGCTGCCGTGCT 201
 RESULT 6
 ID AAZ91119
 XX ID AAZ91119 standard; DNA; 3661 BP.
 XX AC AAZ91119;
 XX DT 06-JUN-2000 (first entry)
 XX DE Mutated Escherichia coli RNase E gene.
 XX KW Ribonuclease E; RNase E; mutation; mRNA degradation; wild type;
 XX KW protein truncation; recombinant gene expression; ds.
 XX OS Escherichia coli.
 XX OS Synthetic.
 XX FH Key
 XX FT Location/Qualifiers
 441..2198


```
Db      168 AGAAAGATCACTGCCCTGGT 147
||||| |||| |||| |||| ||||
RESULT 11
ABA51446/c
ID ABA51446 standard; DNA; 412 BP.
XX
AC ABA51446;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #10141.
XX
KW Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000662.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.
XX
PS Claim 4; SEQ ID NO 10141; 327pp + Sequence Listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a single exon
CC nucleic acid probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;
Query Match 78.2%; Score 17.2; DB 4; Length 412;
Best Local Similarity 86.4%; Pred. No. 1.8e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AGAAAGTATCGTCGCTGGT 22
||||| |||| |||| |||| ||||
Db      168 AGAAAGATCACTGCCCTGGT 147

RESULT 13
AAK43550/c
ID AAK43550 standard; DNA; 412 BP.
XX
AC AAK43550;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human breast cell single exon nucleic acid probe #10141.
XX
KW Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000662.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts.
XX
PS Claim 4; SEQ ID NO 14865; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;
Query Match 78.2%; Score 17.2; DB 4; Length 412;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 AGAAAGTATCGTCGCTGGT 22
||||| |||| |||| |||| ||||
Db      168 AGAAAGATCACTGCCCTGGT 147

RESULT 13
AAK43550/c
ID AAK43550 standard; DNA; 412 BP.
XX
AC AAK43550;
XX
DT 06-NOV-2001 (first entry)
```

```

XX Human bone marrow expressed single exon probe SEQ ID NO: 18107.
DE
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
KW
XX Homo sapiens.
OS
XX WO200157276-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000668.
PF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR
XX 26-MAY-2000; 2000US-0207456P.
PR
XX 30-JUN-2000; 2000US-00608408.
PR
XX 03-AUG-2000; 2000US-00632366.
PR
XX 21-SEP-2000; 2000US-0234687P.
PR
XX 27-SEP-2000; 2000US-0236359P.
PR
XX 04-OCT-2000; 2000GB-00024263.
PR
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488900/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
PT
XX Example 4; SEQ ID NO 18107; 658pp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
CC
XX Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;
SQ
Query Match 78.2%; Score 17.2; DB 4; Length 412;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
QY 1 AGAAGTGCCTGCTGCCGTGGT 22
DB 168 AGAAGAGATCACTGCCCTGGT 147
XX
RESULT 14
AAK17729/c
ID AAK17729 standard; DNA; 412 BP.
XX
XX AAK17729;
AC
XX 05-NOV-2001 (first entry)
DT
XX
XX Human brain expressed single exon probe SEQ ID NO: 17720.
DE
XX
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.
KW
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000667.
XX

```

```

PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-483446/52.
DR
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
PT
XX Example 4; SEQ ID NO 17720; 650pp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention
CC
XX Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;
SQ
Query Match 78.2%; Score 17.2; DB 4; Length 412;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
QY 1 AGAAGTGCCTGCTGCCGTGGT 22
DB 168 AGAAGAGATCACTGCCCTGGT 147
XX
RESULT 15
ABS43195/c
ID ABS43195 standard; DNA; 412 BP.
XX
XX ABS43195;
AC
XX 25-FEB-2003 (first entry)
DT
XX
XX Human liver single exon probe, SEQ ID NO 18185.
DE
XX
XX Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
KW
XX Homo sapiens.
XX
XX WO200157273-A2.
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000664.
PF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR
XX 26-MAY-2000; 2000US-0207456P.
PR
XX 30-JUN-2000; 2000US-00608408.
PR
XX 03-AUG-2000; 2000US-00632366.
PR
XX 21-SEP-2000; 2000US-0234687P.
PR
XX 27-SEP-2000; 2000US-0236359P.
PR
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488998/53.
DR

```

XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
XX Claim 4; SEQ ID NO 18185; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 412 BP; 93 A; 102 C; 117 G; 100 T; 0 U; 0 Other;

Query Match 78.2%; Score 17.2; DB 4; Length 412;
Best Local Similarity 86.4%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AGAAAGTGATCGCTGCCGTGGT 22
||||| ||||| ||||| |||||
Db 168 AGAAAGATGACACTGCCCTGGT 147

Search completed: November 26, 2005, 13:40:50
Job time : 16.1112 secs

```
/db_xref="UniProt/TREMBL:O5T664"
/translations="MRNYSGGPPSVINQNETFANI1KPTVVQQAIAQNGILGDF
I1RYDNRSGSDIGIQLVGFYHYPAPKDPPLPKNVVFLDSSASMGTKLRQTKD
ALFTIILDRPQDRFSIIIGFSNRKIKVMKDLHLSVTPDSIRDGKVIYHMSPTGTDIN
GALQAIRLNLKTVKSGIGSVLSIVLTDGKPTVGETHTLKILNLTREARAGQVC
IPTIGINDVDFLLKSLGLENGLRRVHEEDAGSLQIGFYDEIRTLPLSDIRIDY
PSSVQVATKTLPPNFTNGSEI1IACKLVDRKLDHLHVEVTASNSKKFIILKTDVPR
PQAKGDTGSPRGDGGDTNHERLSYLTKELLSMQSDDEPEKERLQRAQ
ALAVSRFLTPFTSMKLRGVPMDGLEEAHGSAMGPEPVQSVRGATQGPGLLK
KPYQRIKISKTSVDGPHFVDFPLSLRVCNFDGQDGLRLVSDHRDGSVTYNG
ELICAPAPNGHKHQRITITILINKPERSYLEITPSRVILDDGDRLLVPCNSVW
VSGKGLVSNANVTVTIQGSIAFVILHLYKPAFPHLGFYIASEGLSNC
HGLGQFLNADARLTEDPAGPSQNLTHPLLOVGEQPEAVLTVKGHVPVWVKRKIY
NGEQIDCWFPARNNAKLIDGEYKDYLAHFPTGMDTWPGNVQGALKLAALQWQHE
GQ"
```

gene

```
join(complement(50428..50596),complement(39256..39300),
complement(25552..25715),complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/genes="ITIH5"
/locus_tag="RP11-676E7.1-001"
join(complement(50428..50596),complement(39256..39300),
complement(25552..25715),complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:23606..27721))
/genes="ITIH5"
```

mRNA

```
Query Match 100.0%; Score 26; DB 8; Length 174361;
Best Local Similarity 100.0%; Pred. No. 0.24; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 0;
QY 1 GCGACACCAATTGATGCTTTAAGA 26
Db 139649 GCGACACCAATTGATGCTTTAAGA 139674
RESULT 8
AL391687 194563 bp DNA linear HTG 20-OCT-2001
LOCUS Homo sapiens chromosome 10 clone RP11-299D3, 2 unordered pieces.
DEFINITION AL391687
ACCESSION AL391687
VERSION HTG; HTGS PHASE1; HTGS_CANCELLED.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE
1 Almeida J.
Direct Submission
Submitted (16-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 12, 2001 this sequence version replaced gi:13990116.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA299D3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 194262 bases at least Q40
Consensus quality: 194328 bases at least Q30
Consensus quality: 194372 bases at least Q20
Insert size: 194463; sum-of-contigs
Insert size: 173753; 41.5% error; agarose-fp
Quality coverage: 13.33x in Q20 bases; sum-of-contigs Quality
coverage: 14.92x in Q20 bases; agarose-fp
```

AL391687

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CDS

```
match: CDNAs: AB075833.1 AK027375.1 AK075381.1 AL833203.1"
join(complement(50428..50517),complement(39256..39300),
complement(25552..25715),complement(24379..24480),
complement(20853..21103),
complement(AL355374.14:81450..81619),
complement(AL355374.14:80319..80435),
complement(AL355374.14:50238..50406),
complement(AL355374.14:44092..44401),
complement(AL355374.14:40790..41349),
complement(AL355374.14:37479..37532),
complement(AL355374.14:34005..34121),
complement(AL355374.14:30367..30744),
complement(AL355374.14:27378..27721))
W22615.1
```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 172067: contig of 172067 bp in length
* 172068 172167: gap of 100 bp
* 172168 194563: contig of 22396 bp in length.
*
* Location/Qualifiers
* 1..194563
* /organism="Homo sapiens"
* /mol_type="genomic DNA"
* /db_xref="taxon:9606"
* /chromosome="10"
* /clone="RP11-299D3"
* /clone_lib="RPC1-11.2"
* 1..172067
* /note="assembly fragment:00152
* fragment_chain:1"
* 172168..194563
* /note="assembly fragment:03892
* fragment_chain:1
* clone_end:T7
* vector_side:right"

FEATURES
source
1..194563
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-299D3"
/clone_lib="RPC1-11.2"
1..172067
/note="assembly fragment:00152
fragment_chain:1"
172168..194563
/note="assembly fragment:03892
fragment_chain:1
clone_end:T7
vector_side:right"

misc_feature
1..172067
/note="assembly fragment:00152
fragment_chain:1"
172168..194563
/note="assembly fragment:03892
fragment_chain:1
clone_end:T7
vector_side:right"

ORIGIN
Query Match 100.0%; Score 26; DB 14; Length 194563;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
|||||
DB 58795 GCGAACACCAATTGATGCTTTAAGA 58820

RESULT 9
CP000079_01
WPCOMMENT
Sequence split into 12 fragments LOCUS CP000079 Accession CP000079
Fragment Name Begin End
CP000079_00 1 110000
CP000079_01 100001 210000
CP000079_02 200001 310000
CP000079_03 300001 410000
CP000079_04 400001 510000
CP000079_05 500001 610000
CP000079_06 600001 710000
CP000079_07 700001 810000
CP000079_08 800001 910000
CP000079_09 900001 1010000
CP000079_10 1000001 1110000
CP000079_11 1100001 1130447
Continuation (2 of 12) of CP000079 from base 100001 (CP000079 Leishmania major strain F)

Query Match 77.7%; Score 20.2; DB 2; Length 110000;
Best Local Similarity 88.0%; Pred. No. 1.5e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGAACACCAATTGATGCTTTAAGA 26
|||||
DB 87042 CGAACACCAATTGATGCTTTAAGA 87066

RESULT 10
AX003309/c
LOCUS
DEFINITION Sequence 2 from Patent WO9929845.
ACCESSION AX003309
VERSION AX003309.1 GI:9927126

KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Angulo-Mora, J.F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 2 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)

FEATURES
source
1..1102
Location/Qualifiers
/organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"

ORIGIN
Query Match 76.2%; Score 19.8; DB 6; Length 1102;
Best Local Similarity 91.3%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTA 23
|||||
DB 944 GCTAACACCAATTGATGCTTCA 922

RESULT 11
AX003331/c
LOCUS
DEFINITION Sequence 24 from Patent WO9929845.
ACCESSION AX003331
VERSION AX003331.1 GI:9927146
KEYWORDS
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Angulo-Mora, J.F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 9929845-A 24 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE (FR)

FEATURES
source
1..1390
Location/Qualifiers
/organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"

ORIGIN
Query Match 76.2%; Score 19.8; DB 6; Length 1390;
Best Local Similarity 91.3%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTA 23
|||||
DB 1232 GCTAACACCAATTGATGCTTCA 1210

RESULT 12
MKKIN17/c
LOCUS
DEFINITION Mouse kin17 mRNA for kin17 protein.
ACCESSION X58472
VERSION X58472.1 GI:1045209
KEYWORDS KIN17 gene; kin17 protein; zinc-finger protein.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LJNL at: <http://image.llnl.gov>

Series: IRAK Plate: 123 Row: d Column: 16
 This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction, Similarity but not identity to protein.

FEATURES

source
 1..2710
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="mix FVB/N, C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:5026202"
 /tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months old, gross tissue."
 /clone_lib="NCI CGAP_Mam5"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 76.2%; Score 19.8; DB 9; Length 2710;
 Best Local Similarity 91.3%; Pred. No. 2.4e+02;
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGACACCAATTTGATGCTTTA 23

Db 2634 GCTAACCAATTTGATGCTTCA 2612

RESULT 14

CR9311708/c 19113 bp DNA linear BCT 30-JUN-2005
 Streptococcus pneumoniae strain 1095/39 (serotype 36).

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Bentley, S.D., Aanensen, D., Mavroidi, A., Saunders, D., Rabinowitch, E., Collins, M., Donaghy, K., Harris, D., Kallio, M.S., Murphy, L., Quail, M.A., Samuel, G., Skovsted, I.C., Barrell, B.G., Reeves, P., Parkhill, J., and Spratt, B.G.
 Genetic analysis of the capsular biosynthetic locus from all 90 serotypes of *Streptococcus pneumoniae*

JOURNAL

AUTHORS

TITLE

JOURNAL

Submitted (09-DEC-2004) Submitted on behalf of the Pathogen Sequencing Unit, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: sdb@sanger.ac.uk
 NOTE: This sequence was generated from a PCR product representing the region from dexB to aliA and is not necessarily responsible for the expressed capsule serotype. For a detailed description of how CDS products were predicted see the associated publication.

FEATURES

source

1..19113
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /strain="1095/39"
 /db_xref="taxon:1313"
 <1..137
 /genes="dexB"
 /locus_tag="SPC36_0001"
 <1..137
 /gene="dexB"
 /locus_tag="SPC36_0001"
 /codon_start=3
 /transl_table=11
 /protein_id="CAI34356.1"
 /db_xref="GI:68644251"
 /translation="SNEEQDLTVGKVKSVLIENTLAQEVFEKILVPWDAFCVEMTD

gene

CDS

complement(284..1659)
 /gene="tnp"
 /locus_tag="SPC36_0002"
 /pseudo
 complement(join(284..400,400..1659))
 /gene="tnp"
 /locus_tag="SPC36_0002"
 /note="member of homology group 0017 25"
 /pseudo
 /codon_start=1
 /transl_table=11
 /product="putative IS1202 transposase"
 complement(676..1176)
 /gene="tnp"
 /locus_tag="SPC36_0002"
 /note="HMPfam hit to PF00665, Integrase, catalytic region, score 5.4e-19"
 /pseudo
 1835..3280
 /gene="wzg"
 /locus_tag="SPC36_0003"
 1835..3280
 /gene="wzg"
 /locus_tag="SPC36_0003"
 /note="member of homology group 0000 90"
 /codon_start=1
 /transl_table=11
 /product="integral membrane regulatory protein Wzg"
 /protein_id="CAI34358.1"
 /db_xref="GI:68644252"
 /translation="MSRRFKRSRSQKVRNIVLLTYLLVCFLLFLFKNLIAP
 RYLNVTALVLLVGLLIYKAEKFTILVFLSVLSVFLVQVGLTNR
 LNATNSYSEISVAVLADSDIKNTVLTSTVPTGNTDNENIKLADIKSSQNTDLT
 VDQSSYLAAYKSLIAGTKAIVLNSVFENIIESEYDPYASKIKIYTKGFKKVEAP
 KTSKQSNLYVSGIDTYGPISSVSRSQVNIIMVTNRDTKILLTTTPRDAYVPIADG
 GNNQDKLTAGIYGVDSIHTLENLYGVDIYVRLNFTSFLKMLDGLGGVDVNDQ
 EFSALHGHPFPVGNVHLDSEQALGFVRERYSLADGDRGRNQKVIIVAIQKLTST
 EYLVNYSILQGLSDLTNPMTIEMIDLVNTQLESNGGNYKYNQDLKGTGRMDLPSY
 AMPDNLNLYMEVDDSSLAIVKAAIQDVMGR"
 1835..1951
 /gene="wzg"
 /locus_tag="SPC36_0003"
 /note="Signal peptide predicted for SPC1724 by SignalP 2.0
 HMM (Signal peptide probability 0.999) with cleavage site
 probability 0.540 between residues 39 and 40"
 join(1883..1942,1970..2038,2057..2125)
 /gene="wzg"
 /locus_tag="SPC36_0003"
 /note="3 probable transmembrane helices predicted for
 SPC1724 by TMHMM2.0 at aa 17-36, 46-68 and 75-97"
 2042..2386
 /gene="wzg"
 /locus_tag="SPC36_0003"
 /note="HMPfam hit to PF02916, DNA polymerase processivity
 factor, score 2.7e-62"
 2564..3007
 /gene="wzg"
 /locus_tag="SPC36_0003"
 /note="HMPfam hit to PF03816, Cell envelope-related
 transcriptional attenuator, score 3.2e-70"
 3282..4013
 /gene="wzh"
 /locus_tag="SPC36_0004"
 3282..4013
 /gene="wzh"
 /locus_tag="SPC36_0004"
 /EC_number="3.1.3.48"
 /note="member of homology group 0001 90"
 /codon_start=1
 /transl_table=11
 /product="protein-tyrosine phosphatase Wzh"
 /protein_id="CAI34359.1"


```

* 61193 77610: contig of 16418 bp in length
* 77611 77710: gap of unknown length
* 77711 86263: contig of 8553 bp in length.
FEATURES
    source
        1..86263
            /organism="Lotus corniculatus var. japonicus"
            /mol_type="genomic DNA"
            /variety="japonicus"
            /db_xref="taxon:34305"
            /chromosome="1"
            /clone="LjT25J06"
            /clone_lib="LjT library"
            /notes="TAC clone: TM1595, synonym: Lotus japonicus"
        11722..11821
            /estimated_length=unknown
        12922..13021
            /estimated_length=unknown
        16158..16257
            /estimated_length=unknown
        19597..19696
            /estimated_length=unknown
        23701..23800
            /estimated_length=unknown
        29712..29811
            /estimated_length=unknown
        36937..37036
            /estimated_length=unknown
        48213..48312
            /estimated_length=unknown
        61093..61192
            /estimated_length=unknown
        77611..77710
            /estimated_length=unknown
    ORIGIN

```

```

Query Match          76.2%; Score 19.8; DB 14; Length 86263;
Best Local Similarity 91.3%; Pred. No. 2.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      4 AACACCAATTGATGCTTTAAGA 26
      ||||| ||||| ||||| ||||| |||||
DB     10506 AACACGAATTGATGCTTCAGA 10528

```

Search completed: November 26, 2005, 19:11:07
 Job time : 141.192 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 16.6769 Seconds
(without alignments)
10390.557 Million cell updates/sec

Title: US-09-555-529-19
Perfect score: 26
Sequence: 1 gcgaacacacatttgatgctttaaga 26

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues 9993994
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21.*
1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002as.*
7: geneseqn2002bs.*
8: geneseqn2003as.*
9: geneseqn2003bs.*
10: geneseqn2003cs.*
11: geneseqn2003ds.*
12: geneseqn2004as.*
13: geneseqn2004bs.*
14: geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	100.0	26	AX85567	Aax85567 PCR prime
C 2	26	100.0	591	AH34095	Aah34095 Human col
C 3	26	100.0	1002	AX85551	Aax85551 cDNA of a
C 4	26	100.0	1296	AX85549	Aax85549 cDNA of a
5	22	84.6	398	ACF87901	Adf87901 Human SIR
C 6	21.2	81.5	679	ADQ56943	Adq56943 Novel can
C 7	19.8	76.2	1102	AX85550	Aax85550 cDNA of a
C 8	19.8	76.2	1390	AX85570	Aax85570 cDNA of a
C 9	19.8	76.2	1458	AAQ79936	Aaq79936 Murine Ki
10	19.6	75.4	915	AAQ79262	Aaq79262 Helicobac
11	19.6	75.4	1024	ABX66121	Abx66121 Helicobac
12	19.6	75.4	1116	AX14506	Aax14506 H. pylori
C 13	18.8	72.3	4328	AAQ29635	Aaq29635 Arabidops
C 14	18.8	72.3	4505	AAV15448	Aav15448 Arabidops
C 15	18.8	72.3	14113	AAA39651	Aaa39651 Fusion co
C 16	18.8	72.3	14113	AAA90096	Aaa90096 PR-1 prom
C 17	18.8	72.3	14113	AAH25853	Aah25853 PR-1 prom
C 18	18.8	72.3	134499	Aaf22286	Aaf22286 BAC conta
19	18.2	70.0	2253	AAD54900	Aad54900 Arabidops

20	18	69.2	217	3	AAC30472	Aac30472 Human sec
21	18	69.2	407	10	ADF81241	Adf81241 Leukaemia
22	18	69.2	466	10	ADF82611	Adf82611 Leukaemia
23	18	69.2	516	13	ACN56786	Acn56786 Cotton gy
24	18	69.2	528	13	ACN62182	Acn62182 Cotton gy
C 25	18	69.2	559	13	ACN60657	Acn60657 Cotton gy
C 26	18	69.2	559	13	ACN51102	Acn51102 Cotton an
C 27	18	69.2	574	13	ACN56699	Acn56699 Cotton gy
C 28	18	69.2	599	13	ACN62609	Acn62609 Cotton de
C 29	18	69.2	606	13	ACN62525	Acn62525 Cotton de
30	18	69.2	1839	10	ADF09693	Adf09693 Human pap
31	18	69.2	4498	13	ADX64079	Adx64079 Plant ful
32	18	69.2	110000	13	ABD32909_6	Abd32909 (7 of
C 33	18	69.2	158811	12	ADQ19901	Adq19901 Human sof
34	17.6	67.7	309	8	ACF72766	Adf72766 Staphyloc
35	17.6	67.7	561	11	ACL30673	Adl30673 Rice abio
36	17.6	67.7	672	10	ADF01210	Adf01210 Bacterial
C 37	17.6	67.7	697	6	ABK34383	Abk34383 Human cdn
C 38	17.6	67.7	837	8	ACA30257	Aca30257 Prokaryot
C 39	17.6	67.7	943	6	ABV77378	Abv77378 Glucose f
C 40	17.6	67.7	1024	10	ADC86948	Adc86948 Human GPC
C 41	17.6	67.7	1056	8	ACA28525	Aca28525 Prokaryot
42	17.6	67.7	1200	10	ACC60903	Acc60903 Gene sequ
43	17.6	67.7	1200	10	ADK62329	Adk62329 Disease c
44	17.6	67.7	1238	13	ADT18265	Adt18265 Plant cdn
C 45	17.6	67.7	1247	6	ABL40378	Ab140378 Human pho

ALIGNMENTS

RESULT 1
AX85567
ID AX85567 standard; cDNA; 26 BP.
XX
AC AX85567;
XX
DT 07-SEP-1999 (first entry)
XX
DE PCR primer and probe used to detect human kin17 nucleic acids.
XX
KW Human; kin17 protein; cell proliferation; fertility; probe;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJP;
XX
DR WPI; 1999-359999/31.
XX
PT New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.
XX
PS Claim 7; Page 5; 69pp; French.
XX
CC AAX85553-69 represent PCR primers and probes used to detect nucleic acids
CC encoding human kin17 protein. The mammalian kin17 protein is useful for
CC preparing a medicament for controlling cell proliferation or for
CC controlling fertility. The medicaments can also be used to treat
CC hyperproliferative diseases. Fragments between amino acids 55 and 235
CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein

CC are useful for regulating the interaction between proteins and curved
CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
CC proliferation

XX
SQ Sequence 26 BP; 9 A; 5 C; 5 G; 7 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
|||||
DB 1 GCGAACACCAATTGATGCTTTAAGA 26

RESULT 2

AAH34095/c
ID AAH34095 standard; cDNA; 591 BP.

XX
AC AAH34095;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1177.

XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
OS Homo sapiens.

XX
PN WO200122920-A2.

XX
XX 05-APR-2001.

XX
PF 28-SEP-2000; 2000WO-US026524.

XX
PR 29-SEP-1999; 99US-0157137P.

XX
PR 03-NOV-1999; 99US-0163280P.

XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX
XX WPI; 2001-235357/24.
DR P-PSDB; AAG74690.

XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.

XX
PS Claim 1; Page 2977-2978; 9803pp; English.

XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patient's own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922

XX
SQ Sequence 591 BP; 199 A; 95 C; 114 G; 182 T; 0 U; 1 Other;

Query Match 100.0%; Score 26; DB 4; Length 591;
Best Local Similarity 100.0%; Pred. No. 0.057;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
|||||
DB 326 GCGAACACCAATTGATGCTTTAAGA 301

RESULT 3

AAH85551/c
ID AAX85551 standard; cDNA; 1002 BP.

XX
AC AAX85551;

XX
DT 07-SEP-1999 (first entry)

XX
DE cDNA of a gene coding for a truncated human kin17 protein.

XX
KW kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW protein interaction; curved DNA; HIV replication; HIV integration;
KW repair enzyme; ss.

XX
OS Homo sapiens.

XX
PN FR2772046-A1.

XX
PD 11-JUN-1999.

XX
PF 09-DEC-1997; 97FR-00015536.

XX
PR 09-DEC-1997; 97FR-00015536.

XX
PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX
PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX
XX WPI; 1999-359999/31.

XX
XX New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.

XX
PS Claim 5; Page 32; 69pp; French.

XX
CC The present sequence encodes a truncated human kin17 protein with amino
CC acids 129-228 deleted. The mammalian kin17 protein is useful for
CC preparing a medicament for controlling cell proliferation or for
CC controlling fertility. The medicaments can also be used to treat
CC hyperproliferative diseases. Fragments between amino acids 55 and 235
CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein
CC are useful for regulating the interaction between proteins and curved
CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
CC proliferation

XX
SQ Sequence 1002 BP; 343 A; 189 C; 230 G; 240 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 2; Length 1002;
Best Local Similarity 100.0%; Pred. No. 0.062;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
|||||
DB 1002 GCGAACACCAATTGATGCTTTAAGA 977

RESULT 4

AAH85549/c
ID AAX85549 standard; cDNA; 1296 BP.

XX
AC AAX85549;

```
XX 07-SEP-1999 (first entry)
XX cDNA of a gene coding for the human kin17 protein.
XX Human; kin17 protein; cell proliferation; fertility;
XX hyperproliferative disease; protein interaction; curved DNA;
XX HIV replication; HIV integration; repair enzyme; ss.
XX Homo sapiens.
XX PR2772046-A1.
XX 11-JUN-1999.
XX 09-DEC-1997; 97FR-00015536.
XX 09-DEC-1997; 97FR-00015536.
XX (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
XX New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
XX Claim 1; Page 30-31; 69pp; French.
XX The present sequence encodes a human kin17 protein. The mammalian kin17
XX protein is useful for preparing a medicament for controlling cell
XX proliferation or for controlling fertility. The medicaments can also be
XX used to treat hyperproliferative diseases. Fragments between amino acids
XX 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
XX kin17 protein are useful for regulating the interaction between proteins
XX and curved DNA. The fragment can be used to block replication of HIV or
XX its integration into the human genome or to target repair enzymes to
XX curved DNA sites. Expression vectors for kin17 can be used for
XX controlling cell proliferation
XX Sequence 1296 BP; 453 A; 240 C; 310 G; 293 T; 0 U; 0 Other;
Query Match 100.0%; Score 26; DB 2; Length 1296;
Best Local Similarity 100.0%; Pred. No. 0.064;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GCGAACCAATTTGATGCTTTAAGA 26
Db 1296 GCGAACCAATTTGATGCTTTAAGA 1271
RESULT 5
ACF87901
ID ACF87901 standard; DNA; 398 BP.
XX AC
XX ACF87901;
XX 02-JUN-2005 (first entry)
XX Human SIRS/sepsis diagnostic marker DNA fragment 6761.
XX Systemic inflammatory response syndrome; SIRS; antibacterial;
XX immunosuppressive; antiinflammatory; diagnosis; sepsis; ds.
XX Homo sapiens.
XX WO2004087949-A2.
XX 14-OCT-2004.
XX 31-MAR-2004; 2004WO-EP003419.
XX
02-APR-2003; 2003DE-01015031.
08-AUG-2003; 2003DE-01036511.
02-SEP-2003; 2003DE-01040395.
(SIRS-) SIRS LAB GMBH.
Russwurm S, Reinhart K, Saluz H, Straube E, Zipfel PF, Deigner H;
WPI; 2004-748070/73.
In vitro detection of systemic inflammatory response syndrome and related
conditions, for e.g. monitoring progression, comprises detecting abnormal
expression of disease-related genes.
Disclosure; Page; 75pp; German.
The invention relates to a novel method for in vitro detection of
systemic inflammatory response syndrome (SIRS). The method comprises
detecting abnormal expression of disease-related genes, or their
associated peptides. The method of the invention demonstrates
antibacterial, immunosuppressive and antiinflammatory applications and
may be used for early differential diagnosis, monitoring progression,
assessing risk, assessing the likely response to treatment and for post
mortem diagnosis of systemic inflammatory response syndrome, sepsis and
sepsis-like conditions. The recombinant or synthetic nucleic acid
sequences of the invention, or derived proteins or peptides, may be
useful as calibrants in assays for the specified diseases, for evaluating
activity or toxicity in screening for active agents and/or for
preparation of agents for treatment or prevention of the specified
diseases. The current sequence is that of a human SIRS/sepsis diagnostic
marker DNA fragment of the invention. Note: The sequence data for this
patent did not form part of the printed specification, but was obtained
in electronic format directly from WIPO at ftp.wipo.int/pub/published
pct_sequences. Furthermore, a number of arbitrary SEQ ID NO.s are
disclosed within the specification, however, these have not been taken
into account during indexing due to inconsistencies in application and
format
XX Sequence 398 BP; 141 A; 64 C; 62 G; 127 T; 0 U; 4 Other;
Query Match 84.6%; Score 22; DB 13; Length 398;
Best Local Similarity 95.7%; Pred. No. 3.9;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 4 AACACCAATTTGATGCTTTAAGA 26
Db 225 AACACCAATTTGATGCTTTAAGA 247
RESULT 6
ADQ56943/c
ID ADQ56943 standard; DNA; 679 BP.
XX AC
XX ADQ56943;
XX 21-OCT-2004 (first entry)
XX Novel canine microarray-related DNA sequence SeqID8245.
XX canine microarray; drug screening; toxicity assay;
XX environmental pollutant; cellular response; gene expression profile;
XX toxic response; liver necrosis; fatty liver disease;
XX protein adduct formation; hepatitis; dog; ds.
XX Canis familiaris.
XX WO2004063324-A2.
XX 29-JUL-2004.
XX 05-MAY-2003; 2003WO-US013853.
XX 03-MAY-2002; 2002US-0377240P.
XX
```



```

XX CC The present sequence encodes a murine kin17 protein. The mammalian kin17
CC protein is useful for preparing a medicament for controlling cell
CC proliferation or for controlling fertility. The medicaments can also be
CC used to treat hyperproliferative diseases. Fragments between amino acids
CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
CC kin17 protein are useful for regulating the interaction between proteins
CC and curved DNA. The fragment can be used to block replication of HIV or
CC its integration into the human genome or to target repair enzymes to
CC curved DNA sites. Expression vectors for kin17 can be used for
CC controlling cell proliferation
XX SQ Sequence 1390 BP; 482 A; 256 C; 353 G; 299 T; 0 U; 0 Other;
Query Match 76.2%; Score 19.8; DB 2; Length 1390;
Best Local Similarity 91.3%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGAACACCAATTGTGATGCTTTA 23
Db 1232 GCTAACACCAATTGTGATGCTTCA 1210
RESULT 9
AAQ79936/c
ID AAQ79936 standard; cDNA; 1458 BP.
XX AC AAQ79936;
XX AC
XX XX 25-MAR-2003 (revised)
DT 06-SEP-1995 (first entry)
XX XX Murine Kin17 cDNA.
DE XX chromosomal rearrangement; kin17 protein; SOS DNA repair system; RecA;
KW genotoxic agent; zinc finger; DNA binding protein; ss.
KW Mus musculus.
OS XX
XX FH Location/Qualifiers
XX primer_bind complement(1..21)
FT FT /tag= b
FT FT /note= "Oligo L (AAQ79946) binds to complementary strand"
FT FT misc_feature 22..1434
FT FT /tag= 1
FT FT /label= kin17 cDNA
FT FT /note= "nucleotides 1-1414; the genomic DNA contains at
FT FT least two introns within this sequence, see Comments"
FT FT primer_bind 32..49
FT FT /tag= c
FT FT /note= "Oligo C (AAQ79938) binding site"
FT FT CDS 46..1221
FT FT /tag= a
FT FT /product= "Kin17"
FT FT /note= "N's in the sequence denote illegible residues"
FT FT primer_bind complement(67..86)
FT FT /tag= d
FT FT /note= "Oligo S (AAQ79947) binds to complementary strand"
FT FT primer_bind 274..297
FT FT /tag= e
FT FT /note= "Oligo D (AAQ79939) binding site"
FT FT primer_bind complement(339..360)
FT FT /tag= f
FT FT /note= "Oligo K (AAQ79945) binds to complementary strand"
FT FT primer_bind 451..474
FT FT /tag= g
FT FT /note= "Oligo J (AAQ79944) binding site"
FT FT primer_bind complement(550..567)
FT FT /tag= h
FT FT /note= "Oligo E (AAQ79940) binds to complementary strand"
FT FT primer_bind 802..825
FT FT /tag= i
FT FT /note= "Oligo F (AAQ79941) binding site"

```

```

FT primer_bind complement(839..862)
FT /tag= j
FT /note= "Oligo G (AAQ79942) binds to complementary strand"
FT primer_bind complement(1435..1458)
FT /tag= k
FT /note= "Oligo B (AAQ79937) binds to complementary strand"
FT XX PR2706487-A1.
XX XX
XX PD 23-DEC-1994.
XX XX
XX PF 15-JUN-1993; 93FR-00007171.
XX XX
XX PR 15-JUN-1993; 93FR-00007171.
XX XX
XX PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX XX
XX PI Angulo-Mora JF, Tissier A, Frelat G, Mauffrey P, Guilly M;
XX XX WPI; 1995-039031/06.
XX XX
XX XX Purified murine kin17 protein prepn. for detecting chromosomal
XX PT rearrangements - also related antibodies, human and murine DNA, primers,
XX PT probes and vectors, used to assess damage caused by genotoxic agents.
XX PS Claim 9; Page 33; 54pp; French.
XX CC The murine Kin17 protein includes a zinc finger domain (see AAR66766),
XX CC recognises single- and double-stranded DNA (partic. regions of secondary
XX CC structure), has apparent mol. wt. 43 kD and is recognised by both anti-
XX CC Kin17 antibodies and antibodies against the RecA protein of E.coli. The
XX CC Kin17 protein is involved in DNA repair; it can be used to monitor
XX CC chromosomal rearrangements following exposure to genotoxic agents. The
XX CC kin17 cDNA sequence AAQ79936 consists of a 1414 nucleotide sequence,
XX CC flanked by primer binding sites; the genomic kin17 DNA is claimed in
XX CC which an intron is inserted at position 137 of the 1414 nucleotide cDNA
XX CC and a second intron is located between nucleotides 339-429 of the cDNA.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
XX SQ Sequence 1458 BP; 496 A; 272 C; 372 G; 311 T; 0 U; 7 Other;
Query Match 76.2%; Score 19.8; DB 2; Length 1458;
Best Local Similarity 91.3%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGAACACCAATTGTGATGCTTTA 23
Db 1277 GCTAACACCAATTGTGATGCTTCA 1255
RESULT 10
AAC97262
ID AAC97262 standard; DNA; 915 BP.
XX ID
XX AC AAC97262;
XX XX
XX DT 23-FEB-2001 (first entry)
XX XX
XX DE Helicobacter pylori bait polypeptide nucleotide sequence #34.
XX KW Helicobacter pylori; two-hybrid system; protein-protein interaction;
XX KW antibacterial; bait polypeptide; gastric ulcer; ds.
XX OS Helicobacter pylori.
XX XX WO200066722-A1.
XX XX
XX XX 09-NOV-2000.
XX XX
XX PF 14-APR-2000; 2000WO-IB000603.
XX XX
XX PR 30-APR-1999; 99EP-00401066.
XX XX

```

PA (HYBR-) HYBRIGENICS SA.
 XX Legrain P, Selig L, Rain J;
 XX WPI: 2000-687535/67.
 DR P-PSDB; AAB52516.
 XX A two-hybrid system for identifying compounds useful in the treatment of
 PT e.g. gastric ulcers comprises producing a collection of recombinant cell
 PT clones.
 XX
 XX Example 5; Page 118-119; 267pp; English.
 XX
 CC The present sequence encodes a bait polypeptide used in a Helicobacter
 CC pylori two-hybrid screen to identify protein-protein interactions. The
 CC method is used to identify a recombinant cell clone expressing a prey
 CC polypeptide which is capable of interacting with the bait polypeptide.
 CC The two hybrid system is useful for screening compounds for antibacterial
 CC activity. It may be used in the treatment of gastric ulcers. The
 CC polynucleotides are useful as amplification primers or specific detection
 CC probes. The polypeptides, vectors or host cells can be used as immunogens
 CC to produce mono- or polyclonal antibodies. The polynucleotides,
 CC polypeptides, antibodies, vectors, host cells or modulating agents can be
 CC used to produce a pharmaceutical composition
 XX
 SQ Sequence 915 BP; 291 A; 161 C; 216 G; 247 T; 0 U; 0 Other;
 Query Match 75.4%; Score 19.6; DB 3; Length 915;
 Best Local Similarity 84.6%; Pred. No. 59;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GCGAACCAACCAATTGATGCTTTAAGA 26
 Db 536 GCGATCTCAAAATTGATGCTTTAAGA 561
 ||||| - - - - - |||||
 RESULT 11
 ABX66121
 ID ABX66121 standard; DNA; 1024 BP.
 XX
 AC ABX66121;
 XX
 DT 07-MAY-2003 (first entry)
 DE Helicobacter pylori selected interacting domain (SID) DNA #720.
 XX
 KW Protein-protein interaction; ulcer; selected interacting domain; SID;
 KW gene; ds.
 XX
 OS Helicobacter pylori.
 XX
 PN WQ200266501-A2.
 XX
 PD 29-AUG-2002.
 XX
 PF 28-DEC-2001; 2001WO-EP015428.
 XX
 PR 02-JAN-2001; 2001US-0259302P.
 XX
 PA (HYBR-) HYBRIGENICS.
 PA (INSP) INST PASTEUR.
 XX
 XX Legrain P, Rain J, Colland F, De Reuse H, Labigne A;
 XX WPI: 2002-674910/72.
 DR P-PSDB; AB051377.
 XX
 PT New complexes of protein-protein interactions in Helicobacter pylori,
 PT useful for identifying modulating compounds for treating or preventing
 PT ulcers in mammals.
 XX
 PS Claim 7; Page 260; 642pp; English.
 XX

CC The invention describes a complex of protein-protein interactions in
 CC Helicobacter pylori selected from 421 complexes given in the
 CC specification. The complex of protein-protein interactions are useful for
 CC screening for agents which modulate the interaction of proteins.
 CC Modulating compounds which binds to a targeted bacterial protein may be
 CC used for treating or preventing ulcers in a human or animal. This
 CC sequence encodes a selected interacting domain (SID), identified via
 CC protein-protein interactions
 XX
 SQ Sequence 1024 BP; 326 A; 183 C; 233 G; 282 T; 0 U; 0 Other;
 Query Match 75.4%; Score 19.6; DB 6; Length 1024;
 Best Local Similarity 84.6%; Pred. No. 60;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GCGAACCAACCAATTGATGCTTTAAGA 26
 Db 537 GCGATCTCAAAATTGATGCTTTAAGA 562
 ||||| - - - - - |||||
 RESULT 12
 AAX14506
 ID AAX14506 standard; DNA; 1116 BP.
 XX
 AC AAX14506;
 XX
 DT 31-MAR-1999 (first entry)
 XX
 DE H. pylori GHPO 1220 gene.
 XX
 KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease; ss.
 XX
 OS Helicobacter pylori.
 XX
 FH Key Location/Qualifiers
 FT CDS 70..1058
 FT /*tag= a
 XX
 PN WQ9843478-A1.
 XX
 PD 08-OCT-1998.
 XX
 PF 01-APR-1998; 98WO-US006371.
 XX
 PR 01-APR-1997; 97US-00833457.
 PR 24-JUN-1997; 97US-00881227.
 PR 29-JUL-1997; 97US-00902615.
 XX
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Kleanthous H, Al-Garawi A, Miller C, Tomb J, Oomen RP;
 XX
 DR WPI; 1998-542293/46.
 DR P-PSDB; AAW98787.
 XX
 XX New isolated Helicobacter polynucleotides - used to develop products for
 PT the diagnosis, prevention and treatment of Helicobacter infections and
 PT gastrointestinal diseases.
 PT
 XX Claim 1; Page 1713-1714; 2054pp; English.
 XX
 CC This sequence represents a polynucleotide of the invention. It was
 CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
 CC The polypeptides can be used for preventing or treating Helicobacter
 CC infections, and gastroduodenal diseases associated with these infections,
 CC including acute, chronic, and atrophic gastritis, and peptic ulcer for the
 CC diseases, e.g. gastric and duodenal ulcers. They can also be used for the
 CC production of antibodies. The products can also be used for detection and
 CC diagnosis
 XX
 SQ Sequence 1116 BP; 356 A; 203 C; 255 G; 302 T; 0 U; 0 Other;

Query Match 75.4%; Score 19.6; DB 2; Length 1116;
Best Local Similarity 84.6%; Pred. No. 61;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGTGCTTTTAAAG 26
|||||
Db 686 GCGATCTCAAAATTGTGCTTTTAAAG 711

RESULT 13
AAD29635/c
ID AAD29635 standard; DNA; 4328 BP.

XX AC AAD29635;

XX DT 17-MAY-2002 (first entry)

XX DE Arabidopsis thaliana isochorismate synthase related DNA #1.

XX KW Isochorismate synthase; ICS; systemic acquired resistance; SAR; SA;
KW salicylic acid; disease resistance; powdery mildew; pathogenic nematode;
KW tobacco mosaic virus; TMV; tobacco necrosis virus; TNV; gene therapy;
XX antifungal; antibacterial; antiviral; plant; enzyme; ds.

XX OS Arabidopsis thaliana.

XX PN WO200206447-A2.

XX PD 24-JAN-2002.

XX PF 18-JUL-2001; 2001WO-US022636.

XX PR 18-JUL-2000; 2000US-0219231P.

XX PA (GEO) GEN HOSPITAL CORP.

XX PI Wildermuth MC, Ausubel FM, Dewdney J;

XX DR WPI; 2002-179788/23.

XX KW Inducing disease resistance on plant to bacterial and fungal pathogens
e.g. powdery mildew or leaf-spotting bacterial pathogen, comprises
transforming plant cell with nucleic acid molecule encoding isochorismate
synthase.

XX PS Disclosure; Page 108-109; 119pp; English.

XX CC The present invention relates to methods of conferring enhanced disease
resistance on a plant or plant component to a pathogen. The method
involves introducing a nucleic acid molecule encoding isochorismate
synthase (ICS) into a plant cell to yield transformed plant cell and
generating a plant or plant component from transformed plant cell, where
the nucleic acid molecule is expressed in cells of the plant and confers
enhanced disease resistance on the plant. Isochorismate synthases are
useful for synthesizing salicylic acid (SA) which plays an important role
in systemic acquired resistance (SAR). The methods are used to confer
enhanced disease resistance on plants or plant components to a pathogen
including a fungal, especially powdery mildew e.g. Erysiphe orontii or E.
cichoracearum, or a bacterial pathogen, preferably a leaf-spotting
bacterial pathogen such as Xanthomonas, Erwinia and a virulent or
avirulent strain of Pseudomonas syringae. It confers disease resistance
on a plant to pathogenic nematodes including root-knot nematodes (e.g.
Meloidogyne sp.), cyst nematodes (e.g. Heterodera sp.), root-attacking
nematodes (e.g. Rotylenchulus reniformis) and above-ground nematodes
(e.g. Anguina funesta) and viral pathogens including tobacco mosaic virus
(TMV), tobacco necrosis virus (TNV), potato leaf roll virus, potato virus
X, potato virus Y, tomato ring spot virus and tomato spotted wilt virus.
Sequences of the invention are also used for gene therapy. The present
sequence is Arabidopsis thaliana isochorismate synthase (At ICS) related
DNA

XX SQ Sequence 4328 BP; 1456 A; 562 C; 745 G; 1462 T; 0 U; 3 Other;

Query Match 72.3%; Score 18.8; DB 6; Length 4328;
Best Local Similarity 90.9%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AACACCAATTGTGCTTTTAAAG 25
|||||
Db 250 AACACCAATTGTGCTTTTAAAG 229

RESULT 14

AAV15448/c

ID AAV15448 standard; DNA; 4505 BP.

XX AC AAV15448;

XX DT 20-JUL-1998 (first entry)

XX DE Arabidopsis PR-1 promoter.

XX KW Promoter PR-1; salicylic acid, 2,6-dichloroisonicotinic acid;
benzo(1,2,3)thiadiazole-7-carboxylic acid 8-methyl ester;
transgenic plant; ds.

XX OS Arabidopsis thaliana.

XX PH Key Location/Qualifiers

FT promoter 1..4258

FT FT /*tag= a

FT FT /note= "full-length PR-1 promoter"

FT FT 2966..4258

FT FT /*tag= d

FT FT /note= "PR-1 promoter fragment in pLTD6D, used for LS
construct construction"

FT FT 3444..4258

FT FT /*tag= e

FT FT /note= "PR-1 promoter fragment in pLTD7D, confers similar
level of chemical induction as full-length promoter"

FT FT 3554..3563

FT FT /*tag= h

FT FT /note= "LS1"

FT FT 3561..4258

FT FT /*tag= f

FT FT /note= "PR-1 promoter fragment in pLTD7D, confers
chemical induction of gene expression at levels reduced
by approx. 3-4 fold as compared to full-length promoter"

FT FT 3564..3573

FT FT /*tag= i

FT FT /note= "LS2"

FT FT 3574..3583

FT FT /*tag= j

FT FT /note= "LS3"

FT FT 3584..3593

FT FT /*tag= k

FT FT /note= "LS4"

FT FT 3594..3603

FT FT /*tag= l

FT FT /note= "LS5"

FT FT 3604..3613

FT FT /*tag= m

FT FT /note= "LS6"

FT FT 3614..3623

FT FT /*tag= n

FT FT /note= "LS7"

FT FT 3624..3633

FT FT /*tag= o

FT FT /note= "LS8"

FT FT 3634..3643

FT FT /*tag= p

FT FT /note= "LS9"

FT FT 3638..4258

FT FT /*tag= g

FT FT /note= "PR-1 promoter fragment in pLTD72D, confers no

FT primer_bind chemical induction of gene expression"
FT 3644. .3653
FT /*tag= q
FT /note= "LS10"
FT primer_bind 3654. .3663
FT /*tag= f
FT /note= "LS11"
FT primer_bind 3664. .3671
FT /*tag= s
FT /note= "LS12"
FT primer_bind 3672. .3681
FT /*tag= t
FT /note= "LS13"
FT TATA_signal 4229. .4232
FT /*tag= b
FT CDS 4294. .4296
FT /*tag= c
FT /note= "start codon for translation"
XX WO9803536-A1.
XX
XX PD 29-JAN-1998.
XX
XX PF 18-JUL-1997; 97WO-US012626.
XX
XX PR 23-JUL-1996; 96US-0027228P.
XX
XX PA (NOVS) NOVARTIS CORP.
XX
XX PI Lebel EG, Ryals JA, Thorne L, Uknes SJ, Ward ER;
XX
XX DR WPI; 1998-120690/11.
XX
XX PT New chemically inducible promoter from Arabidopsis - used to regulate
XX gene expression in response to e.g. salicylic acid.
XX
XX PS Claim 1; Page 37-39; 60pp; English.
XX
XX CC This DNA sequence comprises the full-length, chemically-inducible
XX Arabidopsis PR-1 promoter in plasmid pUTDID. Cis-acting regulatory
XX elements in the PR-1 promoter involved in chemical induction have been
XX characterised using deletion and linker-scanning mutagenesis and in vivo
XX footprinting. At least a portion of the region of the promoter between
XX positions -698 and -621 (relative to the transcription start site of the
XX PR-1 gene) is required for induction of gene expression by such chemicals
XX as salicylic acid, 2,6-dichloroisonicotinic acid and
XX benzo(1,2,3)thiadiazole 7-carboxylic acid 8-methyl ester. The use of
XX chemically inducible PR-1 promoter fragments to regulate gene expression
XX in plants in the presence of inducing chemicals is disclosed, as well as
XX use of these elements for the isolation of transcriptional regulatory
XX proteins involved in the promoter regulation and for the construction of
XX inducible hybrid promoters. Typical genes which may be regulated are
XX genes that control flowering and fruit ripening, impart tolerance to
XX herbicides or pests, produce enzymes or secondary metabolites, impart
XX sterility or dwarfism, improve nutritional qualities etc., also those
XX CC that express antisense sequences or are involved in production of
XX plastics, e.g. polyhydroxybutyrate, or their precursors, perfumes,
XX pharmaceuticals etc
XX
XX SQ Sequence 4505 BP; 1501 A; 705 C; 796 G; 1500 T; 0 U; 3 Other;

Query Match 72.3%; Score 18.8; DB 2; Length 4505;
Best Local Similarity 90.9%; Pred. No. 1.8e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AACACCAATTGATGCTTTAAG 25
DB 250 AACACCAATTGATGCTTTAAG 229

RESULT 15
AAA39651/C
ID AAA39651 standard; DNA; 14113 BP.

XX AAA39651;
XX AC
XX DT 11-SEP-2000 (first entry)
XX DE Fusion construct PR-1/luc DNA.
XX PR-1; luciferase; firefly; PR-1/luc; plant disease; resistance; mutant;
KW systemic acquired resistance; SAR; lesion mimic phenotype; antibacterial;
KW antiviral; fungicide; plant breeding program; fungicide screening;
KW pathogen; ss.
XX Arabidopsis thaliana.
OS Photinus pyralis.
OS Synthetic.
XX US6057490-A.
XX PN
XX PD 02-MAY-2000.
XX PF 30-DEC-1998; 98US-0023134.
XX PR 08-JAN-1993; 93US-00002285.
PR 10-DEC-1993; 93US-00165238.
PR 16-MAY-1996; 96US-00648949.
PR 18-DEC-1997; 97US-00992801.
XX
XX PA (NOVS) NOVARTIS FINANCE CORP.
XX
XX PI Uknes SJ, Ward ER, Maleck K, Ryals JA;
XX
XX DR WPI; 2000-338627/29.
XX
XX PT Selecting disease resistant mutant plants useful for plant breeding,
XX involves selecting uninfected transformed plants that constitutively
XX express a reporter gene regulated by SAR gene promoter.
XX
XX PS Claim 5; Col 25-38; 20pp; English.
XX
XX CC This invention describes a novel method for selecting disease resistant
XX mutant plants (I) by generating a population of plants transformed with a
XX reporter gene (RG) under the regulation of systemic acquired resistance
XX (SAR) gene promoter and selecting uninfected phenotypically normal plants
XX that are lacking a lesion mimic phenotype and constitutively expressing
XX RG, in the absence of viral, bacterial or fungal infections. The products
XX of the invention have antibacterial, antiviral and fungicidal activity.
XX Selected disease resistant mutant plants are useful in plant breeding
XX programs. They also have utility in disease and pathogenesis testing and
XX fungicide screening programs. This sequence represents a fusion construct
XX PR-1/luc DNA which is constructed from Arabidopsis thaliana PR-1 gene and
XX the firefly (Photinus pyralis) luciferase gene, and is used in the method
XX of the invention
XX SQ Sequence 14113 BP; 3645 A; 3225 C; 3345 G; 3766 T; 0 U; 132 Other;

Query Match 72.3%; Score 18.8; DB 3; Length 14113;
Best Local Similarity 90.9%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AACACCAATTGATGCTTTAAG 25
DB 7972 AACACCAATTGATGCTTTAAG 7951

Search completed: November 26, 2005, 13:40:52
Job time : 18.6769 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:58:53 ; Search time 145.336 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-19

Perfect score: 26

Sequence: 1 gcgaacacccaatttgatgctttaaga 26

Scoring table: IDENTITY_NUC

Gapop 10,0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_hic: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_est7: *
9: gb_gss1: *
10: gb_gss2: *
11: gb_gss3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	249	1	AI493544
2	26	100.0	328	2	BE091046
3	26	100.0	384	1	AA770446
4	26	100.0	384	1	AI493563
5	26	100.0	386	3	BM661987
6	26	100.0	469	1	AI087818
7	26	100.0	476	1	AI0878396
8	26	100.0	485	1	AI089251
9	26	100.0	488	3	BM689918
10	26	100.0	512	5	BQ548883
11	26	100.0	516	2	BG944189
12	26	100.0	521	2	BE349628
13	26	100.0	543	6	CB158644
14	26	100.0	575	3	BM750112
15	26	100.0	591	2	BE896845
16	26	100.0	592	8	CX866354
17	26	100.0	605	6	CB242699
18	26	100.0	660	1	AL597250
19	26	100.0	682	1	AV721396
20	26	100.0	698	1	AI650375
21	26	100.0	701	1	AI807250
22	26	100.0	720	7	CK301009

23	26	100.0	733	5	BX104466
24	26	100.0	848	1	AL558810
25	26	100.0	1339	4	CR618602
26	26	100.0	1396	4	CR595908
27	25	96.2	723	3	BQ186497
28	24.4	93.8	259	6	CF526711
29	24.4	93.8	349	1	AA805923
30	24.4	93.8	558	5	BU737321
31	24.4	93.8	679	1	AM028238
32	24.4	93.8	691	1	AM028686
33	24.4	93.8	703	1	AM035210
34	24.4	93.8	776	2	BF571691
35	24.4	93.8	786	1	AJ819816
36	24.4	93.8	864	1	AM037867
37	24.4	93.8	871	1	AJ819723
38	24.4	93.8	949	3	BM453332
39	23	88.5	477	1	AJ556645
40	23	88.5	624	5	BQ602324
41	23	88.5	663	1	AJ660240
42	23	88.5	682	6	CF795929
43	23	88.5	1543	4	AY609991
44	22.8	87.7	800	1	AM038283
45	22	84.6	398	8	H75516

ALIGNMENTS

RESULT 1
AI493544
LOCUS th36a10.x1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2120346 3',
DEFINITION mRNA sequence.
ACCESSION AI493544
VERSION AI493544.1 GI:4394547
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 249)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapps@mail.nih.gov
Life Technologies catalog #: 11548-013
DNA Sequencing by: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 841 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 246.
Location/Qualifiers
1. 249
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2120346"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Pan1"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site: 1; Salt: Site 2; NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"

ORIGIN
Query Match 100.0%; Score 26; DB 1; Length 249;
Best Local Similarity 100.0%; Pred. No. 0.42;

Matches	26;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	245	CGGAACACCAATTGATGCTTTAAGA	220						
RESULT 3									
AA770446									
LOCUS									
Matches	26;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									
BE091046/c									
LOCUS									
DEFINITION									
ACCESSION									
VERSION									
KEYWORDS									
SOURCE									
ORGANISM									
REFERENCE									
AUTHORS									
TITLE									
FEATURES									
source									
QY	1	CGGAACACCAATTGATGCTTTAAGA	26						
Db	221	CGGAACACCAATTGATGCTTTAAGA	246						
RESULT 2									

Life Technologies catalog #: 11548-013
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 843 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 366.

FEATURES

source
 1. .384
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2120370"
 /tissue_type="adenocarcinoma"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Panel"
 /note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.72 kb. Life Technologies catalog #: 11548-013"

ORIGIN

Query Match 100.0%; Score 26; DB 1; Length 384;
 Best Local Similarity 100.0%; Pred. No. 0.45;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGACACCAATTGATGCTTTAAGA 26
 |||||
 Db 221 GCGACACCAATTGATGCTTTAAGA 246

RESULT 5

BM661987
 LOCUS
 DEFINITION UI-E-CK1-abo-f-09-0-UI.s1 UI-E-CK1 Homo sapiens cDNA clone
 UI-E-CK1-abo-f-09-0-UI 3', mRNA sequence.

ACCESSION BM661987.1 GI:18966162
 VERSION
 KEYWORDS
 SOURCE EST.

ORGANISM

Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.

1 (bases 1 to 386)

REFERENCE
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

PUBMED 8889548

COMMENT Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bento-soares@uiowa.edu

Tissue Procurement: Dr. Gregg Hageman

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research

Genetics (www.resgen.com).

Seq primer: M13 Forward

POLYA=Yes.

FEATURES

source
 1. .386
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-CK1-abo-f-09-0-UI"
 /tissue_type="Retina Foveal and Macular"

/dev stages="adult"
 /lab host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone lib="UI-E-CK1"
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site_1: EcoR I; Site_2: Not I;
 UI-E-CK1 is a normalized cDNA library containing the
 following tissue(s): Retina Foveal and Macular. The
 library was constructed according to Bonaldo, Lennon and
 Soares, Genome Research, 6:791-806, 1996. First strand
 cDNA synthesis was primed with an oligo-dT primer
 containing a Not I site. Double stranded cDNA was ligated
 to an EcoR I adaptor, digested with Not I, and cloned
 directionally into pT7T3-Pac vector. The oligonucleotide
 used to prime the synthesis of first-strand cDNA contains
 a library tag sequence that is located between the Not I
 site and the (dT)18 tail. The sequence tag for this
 library is GTCC. This library was created for the program,
 Gene Discovery in the Visual System, supported by National
 Eye Institute (NEI).
 TAG_TISSUE=Foveal and Macular Retina
 TAG_LIB=UI-E-CK1
 TAG_SEQ=GTCC"

ORIGIN

Query Match 100.0%; Score 26; DB 3; Length 386;
 Best Local Similarity 100.0%; Pred. No. 0.45;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGACACCAATTGATGCTTTAAGA 26
 |||||
 Db 238 GCGACACCAATTGATGCTTTAAGA 263

RESULT 6

AI087818
 LOCUS
 DEFINITION O024a09.x1 Soares NSF F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 IMAGE:1567096 3', mRNA sequence.

ACCESSION AI087818
 VERSION
 KEYWORDS
 SOURCE EST.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.

1 (bases 1 to 469)

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 846 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 416.

Location/Qualifiers

1. .469

source

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:1567096"

/lab host="DH10B"

/clone lib="Soares NSF F8_9W_OT_PA_P_S1"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site 1: Not I; Site 2: Eco RI;

Equal amounts of plasmid DNA from five normalized

libraries were mixed, and ss circles were made in vitro.

Following HAP purification, this DNA was used as tracer in

a subtractive hybridization reaction. The driver was

PCR-amplified cDNAs from pools of 5,000 clones made from

the same 5 libraries. The pools consisted of the following libraries and clones: Soares NBHP pool 1: 309384-310919, 323208-325895 Soares NB2HP pool 1: 145032-147335, 147720-148103, 148872-149255, 15002 - 150407, 151176-152327 Soares NB2HP8-9W pool 1: 758280-760583, 772104-774407 Soares NBHPA pool 1: 304776-306311, 320136-322823, 326280-326663 Soares NBHOT pool 1: 723720-726407, 739080-740999 Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 26; DB 1; Length 469;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACCAATTTGATGCTTTAAGA 26

Db 231 GCGAACCAATTTGATGCTTTAAGA 256

RESULT 7

AI378396

LOCUS tc78f05.x1 Soares NBHMPU_S1 Homo sapiens cDNA clone IMAGE:2070753
DEFINITION 3', mRNA sequence.

ACCESSION AI378396

VERSION AI378396.1 GI:4188249

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 476)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 625 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 442.

Location/Qualifiers

FEATURES

source

1. .476

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2070753"

/tissue_type="Pooled human melanocyte, fetal heart, and

pregnant uterus"

/lab_host="DH10B"

/clone_lib="Soares NBHMPU S1"

/note="Organ: mixed (see below); Vector: pT7T3D-Pac

(Pharmacia) with a modified polylinker; Site 1: Not I;

Site 2: Eco RI; Equal amounts of plasmid DNA from three

normalized libraries (melanocyte 2NBHM, pregnant uterus

NBHPU, and fetal heart NBHH19W) were mixed, and ss circles

were made in vitro. Following HAP purification, this DNA

was used as tracer in a subtractive hybridization

reaction. The driver was PCR-amplified cDNAs from pools of

5,000 clones made from the same 3 libraries. The pools

consisted of I.M.A.G.E. clones 260232-265223,

340488-345479, and 484488-489479."

ORIGIN

Query Match 100.0%; Score 26; DB 1; Length 476;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACCAATTTGATGCTTTAAGA 26

Db 218 GCGAACCAATTTGATGCTTTAAGA 243

RESULT 8

AI089251

LOCUS qB9a1.x1 Soares_pregnant_uterus_NBHPU Homo sapiens cDNA clone

DEFINITION IMAGE:1695740 3', mRNA sequence.

ACCESSION AI089251

VERSION AI089251.1 GI:3428310

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 485)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 442.

Location/Qualifiers

FEATURES

source

1. 485

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:1695740"

/sex="female"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="Soares_pregnant_uterus_NBHPU"

/note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I;

Site 2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dT) primer [5'

AAC TGAAG AATTCGCGCGCCCTTTT TTTT TTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. Library

went through one round of normalization. Library

constructed by M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 26; DB 1; Length 485;

Best Local Similarity 100.0%; Pred. No. 0.47;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACCAATTTGATGCTTTAAGA 26

Db 218 GCGAACCAATTTGATGCTTTAAGA 243

RESULT 9

BM689918/c

LOCUS BM689918

DEFINITION UI-E-CK1-abo-f-09-0-UI.r1 UI-E-CK1 Homo sapiens cDNA clone

ACCESSION UI-E-CK1-abo-f-09-0-UI 5', mRNA sequence.

VERSION BM689918

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

REFERENCE 1 (bases 1 to 488)

AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.

TITLE
discovery

JOURNAL
Genome Res. 6 (9), 791-806 (1996)

PUBLISHED
8889548

COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRP, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES
source
1. 488
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CK1-abo-f-09-0-UI"
/tissue_type="Retina Foveal and Macular"
/dev_stage="adult"
/lab_host="DHI08 (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-CK1"
/notes="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-CK1 is a normalized cDNA library containing the following tissue(s): Retina Foveal and Macular. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTCCT. This library was created for the program, Gene discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN
Query Match 100.0%; Score 26; DB 3; Length 488;
Best Local Similarity 100.0%; Pred. No. 0.47; Mismatches 0; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy
1 GCGAACCAACCAATTGTGATGCTTTAAGA 26
|||||
Db 463 GCGAACCAACCAATTGTGATGCTTTAAGA 438

RESULT 10
BQ548883/c
LOCUS
DEFINITION
ik93a03.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:6027917 5', similar to, TR:O60870 O60870 KIN17 PROTEIN. ;, mRNA sequence.

BQ548883
BQ548883 512 bp mRNA linear EST 17-JUN-2002
BQ548883.1 GI:21433386
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 512)
Gubin, A.N., Njoroge, J.M., Bouffard, G.G. and Miller, J.L.
Gene expression in proliferating human erythroid cells
Genomics 59 (2), 168-177 (1999)
10409428
Contact: Jeffery L. Miller
Laboratory of Chemical Biology
National Institute of Diabetes and Digestive and Kidney Diseases
Building 10, Room 9B17, National Institutes of Health, Bethesda, MD 20892, USA
Tel: 301 402 2373

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBLISHED
COMMENT

TITLE
JOURNAL
COMMENT

Cardenas, M., Gibbons, M., McCann, R., Cole, R., Teagareishvili, R., Williams, T., Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohpc.harvard.edu
Library was constructed by Dr. J. Ferrer In vivo mass-excised to pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40RP from Gibco
High quality sequence stop: 434.
Location/Qualifiers
1. 512
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6027917"
/tissue_type="insulinoma"
/lab_host="DHI08 (phage-resistant)"
/clone_lib="Human insulinoma"
/notes="Organ: pancreas; Vector: pBluescript SK-; Site 1: XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to pBluescript SK- by Dr. H. Inoue following the Washington University protocol (http://genome.wustl.edu/est/lambda_protocol.shtml). Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permutt Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library."

ORIGIN
Query Match 100.0%; Score 26; DB 5; Length 512;
Best Local Similarity 100.0%; Pred. No. 0.47; Mismatches 0; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy
1 GCGAACCAACCAATTGTGATGCTTTAAGA 26
|||||
Db 360 GCGAACCAACCAATTGTGATGCTTTAAGA 335

RESULT 11
BQ944189
LOCUS
DEFINITION
BQ944189 516 bp mRNA linear EST 15-JAN-2003
BQ944189
Homo sapiens cDNA clone ax47e12 random, mRNA sequence.

ACCESSION
BQ944189.1 GI:14343561
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 516)
Gubin, A.N., Njoroge, J.M., Bouffard, G.G. and Miller, J.L.
Gene expression in proliferating human erythroid cells
Genomics 59 (2), 168-177 (1999)
10409428
Contact: Jeffery L. Miller
Laboratory of Chemical Biology
National Institute of Diabetes and Digestive and Kidney Diseases
Building 10, Room 9B17, National Institutes of Health, Bethesda, MD 20892, USA
Tel: 301 402 2373

Fax: 301 435 5148
 Email: jmf@nih.gov
 DNA Sequencing and analyses by National Institutes of Health
 Intramural Sequencing Center (NISC).
 Plate: 47 row: e column: 12
 Seq primer: -21M13 forward primer (ABI).
 Location/Qualifiers

FEATURES

source

```
1..516
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="ax47e12"
/sex="unknown"
/tissue_type="blood"
/cell_type="Erythroid Cells"
/lab_host="Primary Culture of Peripheral Blood Mononuclear Cells"
/dev_stage="progenitor; EPO responsive CD71++++"
/lab_host="SOLR"
/clone_lib="Hembase; Erythroid Progenitor Cells (LCB:ax library)"
/note="Organ: blood; Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: EcoRI; 65,000 proliferating erythroid cells from the Buffy coat of a blood donation were obtained by flow cytometric separation after a 5-day culture period in the presence of erythropoietin. Total RNA was purified from the sorted cell population using Trizol reagent. RNA (0.3 ug) was converted into double stranded cDNA using Clontech's CapFinder cDNA Library Construction Kit (Clontech) according to the manufacturer's protocol and cloned into EcoRI digested Lambda Zap II vector (Stratagene). The phage library was amplified once prior to in vivo excision in SOLR cells. Individual colonies were grown, and the cDNA inserts were sequenced in high throughput (NIH intramural sequencing center http://www.nisc.nih.gov/)."

```

ORIGIN

Query Match 100.0%; Score 26; DB 2; Length 516;
 Best Local Similarity 100.0%; Pred. No. 0.47;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
 |||||
 Db 255 GCGAACACCAATTGATGCTTTAAGA 280

RESULT 12

BE349628
 LOCUS ht58g10.x1 NCI CGAP Mell15 Homo sapiens cDNA clone IMAGE:3150978 3,
 DEFINITION similar to TR:O60870 O60870 KIN17 PROTEIN. ;, mRNA sequence.

ACCESSION BE349628
 VERSION BE349628.1 GI:9261481
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.

REFERENCE 1 (bases 1 to 521)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life

Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 Seq primer: -40UP from Gibco
 High quality sequence stop: 439.
 Location/Qualifiers

FEATURES

source

```
1..521
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3150978"
/tissue_type="malignant melanoma, metastatic to lymph node"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mell15"
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies."

```

ORIGIN

Query Match 100.0%; Score 26; DB 2; Length 521;
 Best Local Similarity 100.0%; Pred. No. 0.47;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
 |||||
 Db 229 GCGAACACCAATTGATGCTTTAAGA 254

RESULT 13

BE158644
 LOCUS CB158644/c 543 bp mRNA linear EST 29-JAN-2003
 DEFINITION X-EST0218017 L18POOLn1 Homo sapiens cDNA clone L18POOLn1-12-F09
 5', mRNA sequence.

ACCESSION CB158644
 VERSION CB158644.1 GI:28143780
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.

REFERENCE 1 (bases 1 to 543)

AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Ch.K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001

JOURNAL Unpublished (2002)

COMMENT Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 12 row: F column: 09

High quality sequence stop: 543.

FEATURES

source

```
1..543
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L18POOLn1-12-F09"
/cell_line="SNU-354+Cho-CK+Choi-CK+HLK-3"
/lab_host="Top10F"
/clone_lib="L18POOLn1"
/note="Organ: Liver; Vector: pT7T3-Pac; Site 1: EcoRI; Site 2: NotI; The library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from harvested cell culture."

```

ORIGIN

Query Match 100.0%; Score 26; DB 6; Length 543;
 Best Local Similarity 100.0%; Pred. No. 0.47; Indels 0; Gaps 0;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
 |||||
 Db 468 GCGAACACCAATTGATGCTTTAAGA 443

RESULT 14
 BM750112/c
 LOCUS 575 bp mRNA linear EST 04-MAR-2002
 DEFINITION K-EST0025570 S5SNU484 Homo sapiens cDNA clone S5SNU484-4-G04 5',
 mRNA sequence.

ACCESSION BM750112
 VERSION GI:19079730
 KEYWORDS EST.
 SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominiidae; Homo.

REFERENCE

1 (bases 1 to 575)
 Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.

21C Frontier Korean EST Project 2001

JOURNAL

Unpublished (2002)

COMMENT

Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongseung@mail.kribb.re.kr

Plate: 4 row: G column: 04

High quality sequence stop: 575.

FEATURES

source

1..575
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S5SNU484-4-G04"
 /sex="M"
 /tissue_type="Stomach"
 /cell_line="Epithelial"
 /cell_host="SNU-484"
 /lab_host="Top10P"
 /clone_lib="S5SNU484"
 /note="Organ: Stomach; Vector: pTZ18RP1; Site 1: EcoRI;
 Site 2: NotI; The poly (A)+ RNA was decapped with tobacco
 acid pyrophosphatase (TAP) and ligated with DNA-RNA linker
 including EcoRI site by treatment of T4 RNA ligase. The
 first strand cDNA was synthesized from oligo dT-selected
 mRNA by priming with dT-tailed vector. The dT-tailed
 vector was adjusted to have about 60nt. The cDNA vector
 was circularized with E. coli DNA ligase after digestion
 of EcoRI which site is also included in vector. An RNA
 strand converted to a DNA strand by Okayama-Berg method.
 The obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10P by electroporation
 method."

ORIGIN

Query Match 100.0%; Score 26; DB 3; Length 575;
 Best Local Similarity 100.0%; Pred. No. 0.48; Indels 0; Gaps 0;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
 |||||
 Db 371 GCGAACACCAATTGATGCTTTAAGA 346

RESULT 15

BE896845/c

LOCUS 591 bp mRNA linear EST 20-OCT-2000

DEFINITION 601437548F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922787 5',
 mRNA sequence.

ACCESSION BE896845

VERSION BE896845

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominiidae; Homo.

1 (bases 1 to 591)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC/DCTD/BTP

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM9757 row: O column: 12

High quality sequence stop: 587.

Location/Qualifiers

1..591

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3922787"

/tissue_type="melanotic melanoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH MGC 72"

/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

Query Match 100.0%; Score 26; DB 2; Length 591;

Best Local Similarity 100.0%; Pred. No. 0.48;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26

|||||

Db 588 GCGAACACCAATTGATGCTTTAAGA 563

Search completed: November 27, 2005, 00:57:55

Job time : 147.336 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:02 ; Search time 5.38437 Seconds
(without alignments)
8583.479 Million cell updates/sec

Title: US-09-555-529-19

Perfect score: 26

Sequence: 1 gcaaacacaaatttgatgctttaaga 26

Scoring table: IDENTITY INUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/1/ina/1 COMB.seq:*

2: /cgn2_6/ptodata/1/ina/5 COMB.seq:*

3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*

4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*

5: /cgn2_6/ptodata/1/ina/H COMB.seq:*

6: /cgn2_6/ptodata/1/ina/PCITUS COMB.seq:*

7: /cgn2_6/ptodata/1/ina/PP COMB.seq:*

8: /cgn2_6/ptodata/1/ina/RE COMB.seq:*

9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.6	75.4	915	3	US-10-012-819-67
C 2	18.8	72.3	14113	3	US-09-223-134-1
C 3	18.8	72.3	14113	3	US-08-992-801-1
C 4	18.8	72.3	14113	3	US-09-223-535-1
C 5	18.4	70.8	1381	3	US-09-621-976-10260
C 6	18	69.2	217	3	US-09-513-999C-34547
7	18	69.2	302	3	US-08-651-155B-42
8	18	69.2	302	3	US-09-194-036B-42
C 9	17.6	67.7	601	3	US-09-949-016-190762
C 10	17.6	67.7	601	3	US-09-949-016-192622
C 11	17.6	67.7	672	3	US-09-543-681A-1495
C 12	17.6	67.7	1899	3	US-08-956-171E-659
C 13	17.6	67.7	1899	3	US-08-781-986A-659
C 14	17.6	67.7	2094	3	US-09-248-796A-1868
C 15	17.6	67.7	15125	3	US-09-949-016-14647
C 16	17.6	67.7	24740	3	US-09-949-016-13528
C 17	17.6	67.7	74962	3	US-09-685-853A-3
C 18	17.6	67.7	118143	3	US-09-949-016-17196
C 19	17.6	67.7	135010	3	US-09-949-016-17234
C 20	17.6	67.7	198632	3	US-09-949-016-12781
C 21	17.6	67.7	198637	3	US-09-949-016-17393
C 22	17.4	66.9	1930	3	US-09-404-390-1
C 23	17.2	66.2	465	3	US-09-328-352-2204
C 24	17.2	66.2	528	3	US-09-248-796A-5746

C 25	17.2	66.2	552	3	US-09-248-796A-1452	Sequence 1452, Ap
26	17.2	66.2	601	3	US-09-949-016-35288	Sequence 35288, A
27	17.2	66.2	601	3	US-09-949-016-35289	Sequence 35289, A
28	17.2	66.2	601	3	US-09-949-016-132710	Sequence 132710, A
29	17.2	66.2	1038	3	US-09-134-000C-1498	Sequence 1498, Ap
30	17.2	66.2	1386	3	US-09-134-001C-984	Sequence 984, Appl
C 31	17.2	66.2	2837	2	US-08-993-228-11	Sequence 11, Appl
C 32	17.2	66.2	2963	3	US-09-949-016-890	Sequence 890, App
C 33	17.2	66.2	8961	3	US-09-949-016-15910	Sequence 15910, A
C 34	17.2	66.2	93364	3	US-09-949-016-14890	Sequence 14890, A
C 35	17.2	66.2	95890	3	US-09-949-016-16412	Sequence 16412, A
C 36	17.2	66.2	194714	3	US-09-949-016-11869	Sequence 11869, A
C 37	17.2	66.2	196714	3	US-09-949-016-15474	Sequence 15474, A
C 38	17.2	66.2	580073	3	US-08-545-528D-1	Sequence 1, Appli
C 39	17	65.4	655	3	US-09-533-559-6241	Sequence 6241, Ap
C 40	17	65.4	735	3	US-09-107-433-1454	Sequence 1454, Ap
C 41	17	65.4	957	3	US-09-583-110-21	Sequence 21, Appl
C 42	17	65.4	1182	3	US-09-489-039A-4766	Sequence 4766, Ap
C 43	17	65.4	8898	3	US-08-961-527-69	Sequence 69, Appl
C 44	17	65.4	24707	3	US-09-740-027-3	Sequence 3, Appli
C 45	17	65.4	24707	3	US-10-274-968-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-10-012-819-67

; Sequence 67, Application US/10012819

; Patent No. 6916615

; GENERAL INFORMATION:

; APPLICANT: Legrain, Luc

; APPLICANT: Selig, Luc

; APPLICANT: Rain, Jean-Christophe

; TITLE OF INVENTION: Collection of prokaryotic DNA for Two-Hybrid System, Helicobacter

; FILE REFERENCE: B5053

; CURRENT APPLICATION NUMBER: US/10/012,819

; CURRENT FILING DATE: 2001-10-30

; PRIOR APPLICATION NUMBER: EP 99401066.8

; PRIOR FILING DATE: 1999-04-30

; NUMBER OF SEQ ID NOS: 278

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 67

; LENGTH: 915

; TYPE: DNA

; ORGANISM: Helicobacter pylori

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(915)

US-10-012-819-67

Query Match 75.4%; Score 19.6; DB 3; Length 915;

Best Local Similarity 84.6%; Pred. No. 15;

Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CGGACACCAATTTGATGCTTTAAGA 26

Db 536 CGGATCTCAAAATTTGATGCTTTAAGA 561

RESULT 2

US-09-223-134-1/C

; Sequence 1, Application US/09223134

; Patent No. 6057490

; GENERAL INFORMATION:

; APPLICANT: Ryals, John

; APPLICANT: Uknes, Scott

; APPLICANT: Ward, Eric

; APPLICANT: Delaney, Terry

; APPLICANT: Lawton, Kay

; APPLICANT: Weymann, Kris

; APPLICANT: Steiner, Henry-York

```

; APPLICANT: Maleck, Klaus
; TITLE OF INVENTION: Method For Breeding Disease Resistance
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 605749artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/223,134
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/992,801
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/165,248
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/002,285
; FILING DATE: 08-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1673/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14113 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "PR-1/luc construct"
; US-09-223-134-1

```

```

Query Match 72.3%; Score 18.8; DB 3; Length 14113;
Best Local Similarity 90.9%; Pred. No. 56;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 4 AACACCAATTGATGCTTTAAG 25
Db 7972 AACACCAATTGATGCTTTAAG 7951

```

```

RESULT 3
US-09-223-801-1/c
; Sequence 1, Application US/08992801
; Patent No. 6107544
; GENERAL INFORMATION:
; APPLICANT: Ryals, John
; APPLICANT: Uknes, Scott
; APPLICANT: Ward, Eric
; APPLICANT: Delaney, Terry
; APPLICANT: Lawton, Kay
; APPLICANT: Weymann, Kris
; APPLICANT: Steiner, Henry-York
; APPLICANT: Maleck, Klaus
; TITLE OF INVENTION: Method For Breeding Disease Resistance
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6107544artis Corporation

```

```

; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/992,801
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/648,949
; FILING DATE: 16-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/165,248
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/002,285
; FILING DATE: 08-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1673/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14113 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "PR-1/luc construct"
; US-08-992-801-1

```

```

Query Match 72.3%; Score 18.8; DB 3; Length 14113;
Best Local Similarity 90.9%; Pred. No. 56;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 4 AACACCAATTGATGCTTTAAG 25
Db 7972 AACACCAATTGATGCTTTAAG 7951

```

```

RESULT 4
US-09-223-535-1/c
; Sequence 1, Application US/09223535
; Patent No. 6232525
; GENERAL INFORMATION:
; APPLICANT: Ryals, John
; APPLICANT: Uknes, Scott
; APPLICANT: Ward, Eric
; APPLICANT: Delaney, Terry
; APPLICANT: Lawton, Kay
; APPLICANT: Weymann, Kris
; APPLICANT: Steiner, Henry-York
; APPLICANT: Maleck, Klaus
; TITLE OF INVENTION: Method For Breeding Disease Resistance
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6232525artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/223,535
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/992,801
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,248
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/002,285
FILING DATE: 08-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1673/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1413 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "PR-1/luc construct"
US-09-223-535-1

Query Match 72.3%; Score 18.8; DB 3; Length 1413;
Best Local Similarity 90.9%; Pred. No. 56;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 4 AACACCAATTGGATGCTTTAAG 25
Db 7972 AACACCAATTGGATGCTTTAAG 7951

RESULT 5
US-09-621-976-10260/c
Sequence 10260, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 10260
LENGTH: 381
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-10260

Query Match 70.8%; Score 18.4; DB 3; Length 381;
Best Local Similarity 95.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 4 AACACCAATTGGATGCTTTA 23
Db 119 AACACCAATTGGATGCTTTA 100

RESULT 6

US-09-513-999C-34547
Sequence 34547, Application US/09513999C
Patent No. 6783961
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclert, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
Patent No. 6783961
FILE REFERENCE: 59 US2,REG
CURRENT APPLICATION NUMBER: US/09/513,999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 34547
LENGTH: 217
TYPE: DNA
ORGANISM: Homo sapiens
US-09-513-999C-34547

Query Match 69.2%; Score 18; DB 3; Length 217;
Best Local Similarity 80.8%; Pred. No. 67;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GCGAACACCAATTGGATGCTTTAAGA 26
Db 82 GCGAACACCAATTGGATGCTTTAAGA 107

RESULT 7
US-08-651-155B-42
Sequence 42, Application US/08651155B
Patent No. 6365401
GENERAL INFORMATION:
APPLICANT: Mahan Dr., Michael J.
APPLICANT: Conner Mr., Christopher P.
APPLICANT: Hiethoff Mr., Douglas M.
TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION
TITLE OF INVENTION: OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
NUMBER OF SEQUENCES: 255
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chrisman, Bynum & Johnson, P.C.
STREET: 1900 Fifteenth Street
CITY: Boulder
STATE: CO
COUNTRY: USA
ZIP: 80302
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,155B
FILING DATE: 17-MAY-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Petersen Mr., Steven C.
REGISTRATION NUMBER: 36,238
REFERENCE/DOCKET NUMBER: 17060.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/546-1300
TELEFAX: 303/449-5426
TELEX: ABA1475
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-651-155B-42

Query Match      69.2%; Score 18; DB 3; Length 302;
Best Local Similarity 80.8%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
||| ||||||| |||||||
Db 191 GCCTCGCTAAATTGATGCTTTAAGA 216

RESULT 8
US-09-194-036B-42
; Sequence 42, Application US/09194036B
; Patent No. 6548246
; GENERAL INFORMATION:
; APPLICANT: Mahan, Michael J.
; Conner, Christopher P.
; Hiethoff, Douglas M.
; TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION
; OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
; INFECTION
; NUMBER OF SEQUENCES: 255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 755 Page Mill Road
; CITY: Mountain View
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/194_036B
; FILING DATE: 17-No. 6548246-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/08208
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: US 08/651,155
; FILING DATE: 1996-05-17
; ATTORNEY/AGENT INFORMATION:
; NAME: Shantanu Basu
; REGISTRATION NUMBER: 43,318
; REFERENCE/DOCKET NUMBER: 220002060601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 813-5995
; TELEFAX: (650) 494-0792
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 302 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: DNA (other)
; SEQUENCE DESCRIPTION: SEQ ID NO: 42:
US-09-194-036B-42

Query Match      69.2%; Score 18; DB 3; Length 302;
Best Local Similarity 80.8%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGAACACCAATTGATGCTTTAAGA 26
||| ||||||| |||||||
Db 191 GCCTCGCTAAATTGATGCTTTAAGA 216
```


; Sequence 1495, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 1495
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Proteus mirabilis
; US-09-543-681A-1495

Query Match 67.7%; Score 17.6; DB 3; Length 672;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCGAACCAATTTGATGCTTTAA 24
| | | | | | | | | | | | | | | | | | | | | |
Db 224 GCGAACACGAGTTGATGTTAA 247
| | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-08-956-171E-659
; Sequence 659, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956.171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 659:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

Query Match 67.7%; Score 17.6; DB 3; Length 672;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCGAACCAATTTGATGCTTTAA 24
| | | | | | | | | | | | | | | | | | | | | |
Db 224 GCGAACACGAGTTGATGTTAA 247
| | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-08-781-986A-659
; Sequence 659, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 659:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-781-986A-659

Query Match 67.7%; Score 17.6; DB 3; Length 1899;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GAACACCAATTTGATGCTTTAAGA 26
| | | | | | | | | | | | | | | | | | | | | |
Db 1302 GAACACCAATTTGATGCTTTAAGA 1325
| | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-09-248-796A-1868
; Sequence 1868, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132

; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 659:
US-08-956-171E-659

Query Match 67.7%; Score 17.6; DB 3; Length 1899;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GAACACCAATTTGATGCTTTAAGA 26
| | | | | | | | | | | | | | | | | | | | | |
Db 1302 GAACACCAATTTGATGCTTTAAGA 1325
| | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-08-781-986A-659
; Sequence 659, Application US/08781986A
; Patent No. 6737248
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 659:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1899 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-781-986A-659

Query Match 67.7%; Score 17.6; DB 3; Length 1899;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 GAACACCAATTTGATGCTTTAAGA 26
| | | | | | | | | | | | | | | | | | | | | |
Db 1302 GAACACCAATTTGATGCTTTAAGA 1325
| | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-09-248-796A-1868
; Sequence 1868, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132

```
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 1868
; LENGTH: 2094
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-1868

Query Match      67.7%; Score 17.6; DB 3; Length 2094;
Best Local Similarity 83.3%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 GAACACCAATTTCATGCTTTAAGA 26
Db      1897 GAACAACAACTAGAGGCTTTAAGA 1920

RESULT 15
US-09-949-016-14647/c
; Sequence 14647, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14647
; LENGTH: 15125
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(15125)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14647

Query Match      67.7%; Score 17.6; DB 3; Length 15125;
Best Local Similarity 83.3%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      3 GAACACCAATTTCATGCTTTAAGA 26
Db      14672 GAACACAGATTTCATGCTTTAATA 14649

Search completed: November 27, 2005, 01:10:49
Job time : 7.38437 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 ; Search time 29,9737 Seconds
(without alignments)
7173.088 Million cell updates/sec

Title: US-09-555-529-19

Perfect score: 26

Sequence: 1 ggaacaccaaatttgatgctttaaga 26

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
C 1	26	100.0	591	5	US-10-106-698-1187
C 2	19.8	76.2	556	7	US-10-424-599-49273
C 3	19.8	76.2	2002	8	US-10-425-115-18276
4	19.6	75.4	915	5	US-10-012-819-67
5	19.6	75.4	954	7	US-10-335-977-1576
6	19.6	75.4	999	7	US-10-335-977-1577
7	19.6	75.4	1116	3	US-09-882-227-439
C 8	19.2	73.8	3186778	5	US-10-027-632-174961
C 9	19.2	73.8	3186778	6	US-10-027-632-174961
C 10	18.8	72.3	578	4	US-09-925-065A-593526
C 11	18.8	72.3	603	4	US-09-925-065A-722502
C 12	18.8	72.3	4328	3	US-09-908-299-4
C 13	18.4	70.8	656	4	US-09-925-065A-679116
C 14	18.2	70.0	548	4	US-09-925-065A-527754
C 15	18.2	70.0	548	4	US-09-925-065A-527755
C 16	18.2	70.0	548	4	US-09-925-065A-527756
17	18.2	70.0	611	7	US-10-424-599-141373
18	18.2	70.0	2253	6	US-10-171-104A-1
19	18	69.2	65	10	US-11-087-804-11
C 20	18	69.2	65	10	US-11-087-804-13
C 21	18	69.2	158	8	US-10-425-115-39464
22	18	69.2	516	7	US-10-021-323-11567
23	18	69.2	528	7	US-10-021-323-16963

C 24	69.2	18	559	7	US-10-021-323-5883	Sequence 5883, Ap
C 25	69.2	18	559	7	US-10-021-323-15438	Sequence 15438, A
C 26	69.2	18	560	4	US-09-925-065A-902359	Sequence 902359, A
C 27	69.2	18	574	7	US-10-021-323-11480	Sequence 11480, A
C 28	69.2	18	599	7	US-10-021-323-17390	Sequence 17390, A
C 29	69.2	18	606	7	US-10-021-323-17306	Sequence 17306, A
C 30	69.2	18	1839	9	US-10-367-057-194	Sequence 194, App
C 31	69.2	18	4498	7	US-10-425-114-34922	Sequence 34922, A
C 32	69.2	18	5134	8	US-10-425-115-106105	Sequence 106105, A
C 33	69.2	18	5134	8	US-10-723-860-2720	Sequence 2720, Ap
C 34	69.2	18	721377	9	US-10-461-862-163	Sequence 163, App
C 35	69.2	18	721377	9	US-10-424-599-21922	Sequence 21922, A
C 36	69.2	18	721377	9	US-10-424-599-21922	Sequence 21922, A
C 37	69.2	18	721377	9	US-10-424-599-21922	Sequence 21922, A
C 38	69.2	18	721377	9	US-10-424-599-21922	Sequence 21922, A
C 39	69.2	18	721377	9	US-10-424-599-21922	Sequence 21922, A
C 40	69.2	18	721377	9	US-10-424-599-21922	Sequence 21922, A
C 41	69.2	18	721377	9	US-10-424-599-21922	Sequence 21922, A
C 42	69.2	18	721377	9	US-10-424-599-21922	Sequence 21922, A
C 43	69.2	18	721377	9	US-10-424-599-21922	Sequence 21922, A
C 44	69.2	18	721377	9	US-10-424-599-21922	Sequence 21922, A
C 45	69.2	18	721377	9	US-10-424-599-21922	Sequence 21922, A

ALIGNMENTS

RESULT 1

US-10-106-698-1187/c
; Sequence 1187, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: Patent in Ver. 3.0
; SEQ ID NO 1187
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (587)..(587)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1187

Query Match 100.0%; Score 26; DB 5; Length 591;

Best Local Similarity 100.0%; Pred. No. 0.25; Mismatches 0; Indels 0; Gaps 0;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCGAACCAACCAATTGATGCTTTAAGA 26

Db 326 GCGAACCAACCAATTGATGCTTTAAGA 301

RESULT 2

US-10-424-599-49273/c
; Sequence 49273, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei

```
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 49273
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_15500C.1
US-10-424-599-49273

Query Match          76.2%; Score 19.8; DB 7; Length 556;
Best Local Similarity 91.3%; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCGAACCAATTGATGCTTTA 23
Db 172 GCGTACCAATTGATGCTTCA 150

RESULT 3
US-10-425-115-18276
; Sequence 18276, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 18276
; LENGTH: 2002
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_116668C.1
US-10-425-115-18276

Query Match          76.2%; Score 19.8; DB 8; Length 2002;
Best Local Similarity 91.3%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 AACACCAATTGATGCTTTAAGA 26
Db 175 AACACCAATTGATGCTTCTGA 197

RESULT 4
US-10-012-819-67
; Sequence 67, Application US/10012819
; Publication No. US20030017478A1
; GENERAL INFORMATION:
; APPLICANT: Legrain, Pierre
; APPLICANT: Sellig, Luc
; APPLICANT: Rain, Jean-Christophe
; TITLE OF INVENTION: Collection of Prokaryotic DNA for Two-Hybrid System, Helicobacter
; TITLE OF INVENTION: pylori Protein-Protein Interactions and Applications thereof
; FILE REFERENCE: B5053
; CURRENT APPLICATION NUMBER: US/10/012,819
; CURRENT FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: EP 99401066.8
; PRIOR FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 278
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 67
```

```
; LENGTH: 915
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(915)
US-10-012-819-67

Query Match          75.4%; Score 19.6; DB 5; Length 915;
Best Local Similarity 84.6%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCGAACCAATTGATGCTTTAAGA 26
Db 536 GCGATCTCAAAATGATGCTTTAAGA 561

RESULT 5
US-10-335-977-1576
; Sequence 1576, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 1576:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 954 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3) LOCATION 1...954
; SEQUENCE DESCRIPTION: SEQ ID NO: 1576:
US-10-335-977-1576

Query Match          75.4%; Score 19.6; DB 7; Length 954;
Best Local Similarity 84.6%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

Qy 1 GCGAACACCAATTGATGCTTTAAGA 26
|||||
Db 686 GCGATCTCAAAATTGATGCTTTAAGA 711

RESULT 6
US-10-335-977-1577
; Sequence 1577, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragoras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 1577:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 998 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...999
; SEQUENCE DESCRIPTION: SEQ ID NO: 1577:
US-10-335-977-1577
Query Match 75.4%; Score 19.6; DB 7; Length 999;
Best Local Similarity 84.6%; Pred. No. 2e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCGAACACCAATTGATGCTTTAAGA 26
|||||
Db 617 GCGATCTCAAAATTGATGCTTTAAGA 642

RESULT 7
US-09-882-227-439
; Sequence 439, Application US/09882227
; Publication No. US20030158396A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal

; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Coomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20030158396A1a1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/047002
; CURRENT APPLICATION NUMBER: US/09/882,227
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/902,615
; PRIOR FILING DATE: 1997-07-29
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 439
; LENGTH: 1116
; TYPE: DNA
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (70) ... (1065)
US-09-882-227-439
Query Match 75.4%; Score 19.6; DB 3; Length 1116;
Best Local Similarity 84.6%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCGAACACCAATTGATGCTTTAAGA 26
|||||
Db 686 GCGATCTCAAAATTGATGCTTTAAGA 711

RESULT 8
US-10-027-632-174961/C
; Sequence 174961, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 174961
; LENGTH: 3186778
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) ... (3186778)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174961
Query Match 73.8%; Score 19.2; DB 5; Length 3186778;
Best Local Similarity 87.5%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 GAACACCAATTGATGCTTTAAGA 26

```
|||||
Db 2232992 GAACACCAATTAGATGCTTGGAGA 2232969

RESULT 9
US-10-027-632-174961/c
; Sequence 174961, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-04-20
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-11-23
; PRIOR FILING DATE: 1999-11-23
; PRIOR FILING DATE: 1999-09-28
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 174961
; LENGTH: 3186778
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(3186778)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174961

Query Match 73.8%; Score 19.2; DB 6; Length 3186778;
Best Local Similarity 87.5%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GAACACCAATTGATGCTTTAAGA 26
|||||
Db 2232992 GAACACCAATTAGATGCTTGGAGA 2232969

RESULT 10
US-09-925-065A-593526/c
; Sequence 593526, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2000-10-24
; PRIOR FILING DATE: 2000-10-24
; PRIOR FILING DATE: 2000-11-20
; PRIOR FILING DATE: 2000-11-30
; PRIOR FILING DATE: 2001-01-16
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 593526
; LENGTH: 578
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(578)
; OTHER INFORMATION: n = A,T,C or G
US-09-925-065A-593526

Query Match 72.3%; Score 18.8; DB 4; Length 578;
Best Local Similarity 90.9%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ACACCAATTGATGCTTTAAGA 26
|||||
Db 436 ACACAAATTTAATGCTTTAAGA 415

Query Match 72.3%; Score 18.8; DB 4; Length 578;
Best Local Similarity 90.9%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ACACCAATTGATGCTTTAAGA 26
|||||
Db 436 ACACAAATTTAATGCTTTAAGA 415

RESULT 11
US-09-925-065A-722502/c
; Sequence 722502, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2000-10-24
; PRIOR FILING DATE: 2000-10-24
; PRIOR FILING DATE: 2000-11-20
; PRIOR FILING DATE: 2000-11-20
; PRIOR FILING DATE: 2000-11-30
; PRIOR FILING DATE: 2001-01-16
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 722502
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-722502

Query Match 72.3%; Score 18.8; DB 4; Length 603;
Best Local Similarity 90.9%; Pred. No. 4.2e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ACACCAATTGATGCTTTAAGA 26
|||||
Db 94 ACACAAATTTAATGCTTTAAGA 73

RESULT 12
US-09-908-299-4/c
; Sequence 4, Application US/09908299
; Publication No. US20030051273A1
; GENERAL INFORMATION:
; APPLICANT: MARY WILDERMUTH,
; APPLICANT: JULIA DEWDNEY
; TITLE OF INVENTION: SALICYLIC ACID BIOSYNTHETIC GENES AND
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 00786/391002
; CURRENT APPLICATION NUMBER: US/09/908,299
; CURRENT FILING DATE: 2001-07-18
; PRIOR FILING DATE: 2000-07-18
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 4328
; TYPE: DNA
US-09-908-299-4/c
```

```
; ORGANISM: Arabidopsis thaliana
US-09-908-299-4

Query Match      72.3%; Score 18.8; DB 3; Length 4328;
Best Local Similarity 90.9%; Pred. No. 6.2e+02;
Matches 20; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      4 AACACCAATTGATGCTTTAAG 25
      |||||
Db      250 AACACCAATTGATGCTTTAAG 229

RESULT 13
US-09-925-065A-679116
; Sequence 679116, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 679116
; LENGTH: 656
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-679116

Query Match      70.8%; Score 18.4; DB 4; Length 656;
Best Local Similarity 95.0%; Pred. No. 6.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 AACACCAATTGATGCTTTA 23
      |||||
Db      127 AACACCAATTGATGCTTTA 146

RESULT 14
US-09-925-065A-527754/c
; Sequence 527754, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 527754
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-527755/c
; Sequence 527755, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 527755
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-527755

Query Match      70.0%; Score 18.2; DB 4; Length 548;
Best Local Similarity 87.0%; Pred. No. 7.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 GAACACCAATTGATGCTTTAAG 25
      |||||
Db      322 GAGCACCAATATGATGCTTTAAG 300

RESULT 15
US-09-925-065A-527755/c
; Sequence 527755, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 527755
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-527755

Query Match      70.0%; Score 18.2; DB 4; Length 548;
Best Local Similarity 87.0%; Pred. No. 7.5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 GAACACCAATTGATGCTTTAAG 25
      |||||
Db      322 GAGCACCAATATGATGCTTTAAG 300

Search completed: November 27, 2005, 02:22:42
Job time : 37.9737 secs
```

This Page Blank (uspic)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:41:08 ; Search time 9.17647 Seconds
(without alignments)
421.056 Million cell updates/sec

Title: US-09-555-529-19

Perfect score: 26

Sequence: 1 gcaacacacaatttgatgctttaaga 26

Scoring table: IDENTITY|NUC

Gapop 10|0 , Gapext 1.0

Searched: 3205263 seqs, 74304013 residues

Total number of hits satisfying chosen parameters: 6410526

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New:

- 1: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq2.*
- 9: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq3.*
- 10: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	16.8	64.6	1688	1	US-10-510-386-157
C 2	16.2	62.3	423	1	US-10-793-626-2609
C 3	16.2	62.3	579	1	US-10-793-626-1999
C 4	16.2	62.3	579	1	US-10-793-626-2303
C 5	16.2	62.3	1809	1	US-10-793-626-2361
C 6	16.2	62.3	2085	1	US-10-485-517-391
C 7	16.2	62.3	2876	1	US-10-793-626-3778
C 8	16.2	62.3	3169	1	US-10-793-626-3865
C 9	16.2	62.3	3591	1	US-10-793-626-4033
C 10	16.2	62.3	3627	1	US-10-793-626-4311
C 11	16.2	62.3	4116	1	US-10-793-626-4127
C 12	16.2	62.3	40439	1	US-10-993-509-1
C 13	16.1	61.5	1948	1	US-10-510-386-243
C 14	16.1	61.5	3069	1	US-10-793-626-4068
C 15	15.8	60.8	19	9	US-11-101-244-36671
C 16	15.8	60.8	19	9	US-11-083-784-36671
C 17	15.6	60.0	1118	9	US-11-082-389-131
C 18	15.6	60.0	1560	9	US-11-082-389-129
C 19	15.6	60.0	1929	9	US-11-074-176-317
C 20	15.6	60.0	2001	7	US-11-074-176-63
C 21	15.6	60.0	2409	1	US-10-689-742-73
C 22	15.6	60.0	2734	1	US-10-510-386-5
C 23	15.6	60.0	7117	7	US-11-082-544-34

Sequence 3651, Ap
Sequence 351, App
Sequence 2029, Ap
Sequence 200, App
Sequence 347, Appl
Sequence 35, Appl
Sequence 105, App
Sequence 34, Appl
Sequence 3383, Ap
Sequence 346, Ap
Sequence 3644, Ap
Sequence 3392, Ap
Sequence 451, App
Sequence 30, Appl
Sequence 36654, A
Sequence 36679, A
Sequence 605665, A
Sequence 1426752, A
Sequence 36654, A
Sequence 36679, A
Sequence 605665, A

ALIGNMENTS

RESULT 1
US-10-510-386-157/c
; Sequence 157, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294, 204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 157
; LENGTH: 1688
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (501)..(1652)
US-10-510-386-157

Query Match 64.6%; Score 16.8; DB 1; Length 1688;
Best Local Similarity 90.0%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 GAACACCAATTGATGCTTT 22
|||||
Db 443 GAACACCAATTGATGATT 424

RESULT 2
US-10-793-626-2609
; Sequence 2609, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09

```

; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2609
; LENGTH: 423
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2609

Query Match          62.3%; Score 16.2; DB 1; Length 423;
Best Local Similarity 85.7%; Pred. No. 54;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGAACCAATTGATGCTTT 22
   ||||| ||||| ||||| |||||
Db 381 CGAACCAAAATTAGATGCTTT 401

RESULT 3
US-10-793-626-1999/c
; Sequence 1999, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1999
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1999

Query Match          62.3%; Score 16.2; DB 1; Length 579;
Best Local Similarity 85.7%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 CACCAATTTGATGCTTTAAGA 26
   || ||||| ||||| ||||| |||||
Db 94 CATCAATTTGATGATTTAAAA 74

RESULT 4
US-10-793-626-2303/c
; Sequence 2303, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2303
; LENGTH: 579
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2303

Query Match          62.3%; Score 16.2; DB 1; Length 579;
Best Local Similarity 85.7%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 CACCAATTTGATGCTTTAAGA 26
   || ||||| ||||| ||||| |||||
Db 94 CATCAATTTGATGATTTAAAA 74

RESULT 5
US-10-793-626-2361
; Sequence 2361, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2361
; LENGTH: 1809
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-2361

Query Match          62.3%; Score 16.2; DB 1; Length 1809;
Best Local Similarity 85.7%; Pred. No. 74;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CGAACCAATTTGATGCTTT 22
   ||||| ||||| ||||| |||||
Db 1767 CGAACCAAAATTAGATGCTTT 1787

RESULT 6
US-10-485-517-391
; Sequence 391, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 391
; LENGTH: 2085
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-10-485-517-391

Query Match          62.3%; Score 16.2; DB 1; Length 2085;
Best Local Similarity 85.7%; Pred. No. 76;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 CACCAATTTGATGCTTTAAGA 26

```

RESULT 9
US-10-793-626-4033/c
; Sequence 4033, Application US/10793626
; Publication NO. US20050255478A1
; GENERAL INFORMATION:

```

1 PUBLICATION NO.: US20030235476A1
2
3 GENERAL INFORMATION:
4
5 APPLICANT: KIMMERLY, WILLIAM JOHN
6
7 TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
8
9 FILE REFERENCE: PU3480US
10
11 CURRENT APPLICATION NUMBER: US/10/793,626
12
13 CURRENT FILING DATE: 2004-03-04
14
15 PRIOR APPLICATION NUMBER: 60/164,258
16
17 PRIOR FILING DATE: 1999-11-09
18
19 NUMBER OF SEQ ID NOS: 4472
20
21 SOFTWARE: PatentIn Ver. 2.1

```

```
; SEQ ID NO 4127
; LENGTH: 4116
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-4127

Query Match          62.3%; Score 16.2; DB 1; Length 4116;
Best Local Similarity 85.7%; Pred. No. 88;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      6 CACCAATTGATGCTTTAAGA 26
Db      3055 CATCAATTGATGATTTAAAA 3035

RESULT 12
US-10-993-509-1
; Sequence 1, Application US/10993509
; Publication No. US20050250121A1
; GENERAL INFORMATION:
; APPLICANT: Aerisens, Jeroen
; APPLICANT: Athanasios, Maria
; APPLICANT: Brain, Carlos
; APPLICANT: Cohen, Nadine
; APPLICANT: Dain, Bradley
; APPLICANT: Denton, R. Rex
; APPLICANT: Judson, Richard S.
; APPLICANT: Ozdemir, Vural
; APPLICANT: Reed, Carol R.
; TITLE OF INVENTION: NTRK2 Genetic Markers Associated with Progression of Alzheimer's
; FILE OF INVENTION: Disease
; FILE REFERENCE: 2300.0060001
; CURRENT APPLICATION NUMBER: US/10/993,509
; CURRENT FILING DATE: 2004-11-22
; PRIOR APPLICATION NUMBER: US 60/524,637
; PRIOR FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 40439
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2402)..(2402)
; OTHER INFORMATION: n is 'c' or 'g'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2722)..(2722)
; OTHER INFORMATION: n is 'c' or 't'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2799)..(2799)
; OTHER INFORMATION: n is 'g' or 'a'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (3799)..(3898)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (6337)..(6436)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (11729)..(11828)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (14905)..(15004)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
```

```
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (17523)..(17622)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (19842)..(19941)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (22316)..(22415)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (24686)..(24785)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (27030)..(27129)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29535)..(29634)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (31929)..(32028)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (34403)..(34502)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36929)..(37028)
; OTHER INFORMATION: n is 'a', 't', 'g', or 'c'
US-10-993-509-1

Query Match          62.3%; Score 16.2; DB 1; Length 40439;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4 AACACCAATTGATGCTTTAA 24
Db      14214 AAAGCAATATGATGCTTTAA 14234

RESULT 13
US-10-510-386-243/c
; Sequence 243, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 243
; LENGTH: 1948
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (501)..(1448)
US-10-510-386-243

Query Match          61.5%; Score 16; DB 1; Length 1948;
```

Best Local Similarity 79.2%; Pred. No. 93;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 2 CGAACACCAATTGATGCTTTAAG 25
Db 792 CAATCACCGAATTAATGCTGTAAG 769
Search completed: November 27, 2005, 02:44:31
Job time : 10.1765 secs

Db 1 AACACUAAUUUGCUGCUUU 19

RESULT 14
US-10-793-626-4068/c
; Sequence 4068, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4068
; LENGTH: 3069
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4068

Query Match 61.5%; Score 16; DB 1; Length 3069;
Best Local Similarity 79.2%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 GAACACCAATTGATGCTTTAAGA 26
Db 3020 GAACACCAATTGAGAAATTAAAA 2997

RESULT 15
US-11-101-244-36671
; Sequence 36671, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 36671
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-36671

Query Match 60.8%; Score 15.8; DB 8; Length 19;
Best Local Similarity 52.6%; Pred. No. 44;
Matches 10; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

Qy 4 AACACCAATTGATGCTTT 22

This Page Blank (uspto)

C:Superfamily: caldesmon
C:Keywords: actin binding; calmodulin binding; phosphoprotein
F:342,427/Binding site: phosphate (Ser) (covalent) #status predicted
F:433,456/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 7.2%; Score 145.5; DB 1; Length 517;
Best Local Similarity 23.6%; Pred. No. 0.16;
Matches 68; Conservative 46; Mismatches 91; Indels 83; Gaps 12;

QY 48 ESHQQLLASNPQFMDYFSEEFNDFLELLRRFGTKRVHNNIVVNEYISHREHIHM 107
DB 90 ERRQRLQALERQKEFDFTIDGSL-----VPSRREVNVEINEITGKEKV-- 138

QY 108 NATQWETUTDFYKWLGRGLCKVDETPGWYQYIDRDPETIR-----RQLELEKKK 159
DB 139 -----ET-----RQRCBIET-----ETVTKGYQRNMWRQDGEEGK 171

QY 160 KODLDEKTAKFIEEQVRRGLEGEQETPVFTLSRENEBEKVTFNLNKA-----GG 213
DB 172 KEEKDEBEKPEVPTENQVKDNKVE-----KAPKEEMKSVNDRKRGVPEQKAQNG 224

QY 214 SAGATTSK-----SSSLGPSALKLGSAAAG--KRKSSQSSA---QPAKKKKK 264
DB 225 ERELTPKLSFENAFGRSLNKAANAAGSEKLEKQQAVALDELKCRREERRKILE 284

QY 265 LBEKKRTARTDAMLQPGIVVKKIYKLGKGEYHKKGVVKEVIDRYTA 312
DB 285 EEEQKKQFEAE-----RKIREBEKKR--MKEBIERRA 317

RESULT 8
H96770
protein heat shock protein F1017.8 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H96770
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Pederspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
C:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
C:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96770
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-630 <STO>
A:Cross-references: UNIPROT:Q9C911; UNIPARC:UPI00000AAAE1; GB:AE005173; NID:G6939229; PI
C:Genetics:
A:Gene: F1017.8
A:Map position: 1

Query Match 6.6%; Score 134.5; DB 2; Length 630;
Best Local Similarity 20.9%; Pred. No. 0.96;
Matches 73; Conservative 65; Mismatches 121; Indels 91; Gaps 15;

QY 27 YCQMCKQCRDENGPKCHMSHQRQLLASNPQFMDYFSEEFNDFLELLRRFGT 86
DB 309 YCIVCSKFKSEKQKNHQSQKKHKEV---AELRESFTDYEEENEEDIDGLDSPSV 365

QY 87 KRHHNNIVVNEYISHREHIHMNATQWETUTDFYKWLGRGLCKVDETPGWYI----- 139
DB 366 BELHKLQELNIDNE-----RDVKVEVGEADETDEYFVASEDMQG 409

QY 140 --QYIDRPE--TIRQLELEKKKKQ-----DLDDDEKTAKEFIE----EQVRR 179
DB 410 SSESEDEDDMTLLKMWGSKQKQKNNVSVKKEDETEVEIEGDTAEPSFDNQKSTGR 469

QY 180 GLEGKEQETPVFTLSRENEBEKVTFNLN-----GNAGG-----SAGATTSKSSSLGPS 228

Db 470 NKEAKE-----ERNKONAGNDWADDTSKVQIPGEGNPDENMMNATESASGALADSQKD 522
QY 229 ---ALKLGSAAAGKRKES-----SQSSAAPAKKKKSALDEI-----MELEBEKK--RTA 273
DB 523 EANSMEYDNRKSTGRRRSKKGDKNNQELNEKSSEADDTQYVNRDMESQDYKKAPRSK 582

QY 274 RTDAWLQPGIVVKIITKK-----LGEKYHKKKGVVKEVIDRYTAVVK 315
DB 583 KSTR-----GMKTKGTGTTKNSNECDRCGEFESRPTKLHKLHADSGHATVK 628

RESULT 9
T29340
hypotheetical protein F21C10.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T29340
R:Du, Z.; Gattung, S.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F21C10.
A:Reference number: Z20610
A:Accession: T29340
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2541 <DUZ>
A:Cross-references: UNIPROT:Q19663; UNIPARC:UPI0000076CB3; EMBL:U55364; PIDN:AAA97973.1;
A:Experimental source: strain Bristol N2; clone F21C10
C:Genetics:
A:Gene: CESP:F21C10.7
A:Map position: 5
A:Introns: 50/3; 91/3; 153/3; 184/3; 215/1; 254/3; 350/1; 392/3; 538/1; 589/3; 625/3; 751/3

Query Match 6.5%; Score 132.5; DB 2; Length 2541;
Best Local Similarity 21.8%; Pred. No. 6.9;
Matches 72; Conservative 55; Mismatches 124; Indels 79; Gaps 11;

QY 97 EYISHREHHNATQWETLT--DFYKWLGRGLCKVDETPGKW-----YIQVIDRDP--ET 148
DB 852 EFVHKMEVEHVQEQLKHTLVELEBEKLEQQLVKVTEHIQMWQEMIEIIRKMFDPSTPMKT 911

QY 149 IRRQLLEKKKKQDLDDEKTAKEFTEQVRRGLEGEQETPVFTLSRENEBEKV----- 203
DB 912 VQESQELQEKVKLKEALEVQPKIEDV-----AKADIAVKTQISRAVEKQKIRENA 966

QY 204 -----TFNLNKGAGSAGATTSSSLGPSALKLGSAAAGKRKSSQSSAQPAK 253
DB 967 ETLEKKAISATFELSKPQQTIEETKTIELTSL-----EQQTILTPEE 1009

QY 254 KKSALDIMELEBEKKKTARTDAWLQPGI--VVKII--TKKLGKYYHKKGVVKEVIDRY 310
DB 1010 KQ-----ELQILKKTIVEBIQMWQETIEIILQVDKTPKTVQSEETKIKKKNBIQQTV 1061

QY 311 TAVKMTDSGRLLKLDQTHLETVPAPGKRVLVNGGYRGNEGTLSESINEKAFSATIVIE 370
DB 1062 VAQTRLEDSARFRKDEFTTKTVQETMTKQQV-----QQLVKELHERAASQNVVR 1113

QY 371 TGPLKGRV-----EGIQYE 385
DB 1114 QKEERSRVQAPQIITQLKDDVDVDEGCRYE 1143

RESULT 10
T05523
hypotheetical protein F13M23.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C:Accession: T05523
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.;
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z15419
A:Accession: T05523
A:Molecule type: DNA

A;Residues: 1-375 <BEV>	
A;Cross-references: UNIPROT:Q8SW24; UNIPARC:UPI00000A6F0F; EMBL:AL035523	
A;Experimental source: cultivar Columbia; BAC clone F13M23	
C;Genetics:	
A;Map position: 4	
A;Note: F13M23.160	
Query Match	
Best Local Similarity 23.4%; Pred. No. 0.83;	
Matches 62; Conservative 51; Mismatches 102; Indels 50; Gaps 12;	
QY	132 ETPLKGYIQLVDRDPETIRRLQLEKKKKQDLDEBKTAKFIEQVRRGLEG-----KE 185
DB	80 ETPSG-----DGSNNIYGLNL-RQVENVEKPEPEVKMLKSMRTLESIPDAPEKE 132
QY	186 --QETPV--FTE-----LSRENEEEKVTNFKNGAGSGAGATTSSKSSLG 226
DB	133 DPESFPVDGFEALLAGYWKPGQIGLKAKEDVKIVEKWSGNEGFGFGKSDKAMKMI 192
QY	227 PSALKLLGSNLS-----GKRKSSQSSAQPAKKKKSALD-----EIMELEE---EKKRT 272
DB	193 DN--KLVGSGSHEVKVGINKEIMBKDRVVRKRNRETEGESTEVKAACKQNTVRGQTR 250
QY	273 ARTDAMLQPGIVVYKIITKKL-GEKYHKKKGVVKEVIDRYTAVVVKMTDSGDRKL-LDQTHL 330
DB	251 REKTSWLRSHIKVRIISKDVKGRLYKKAIVTDVVGPTSCDIAMDETQLVQGIQELL 310
QY	331 ETVIPAPGRVVLVNGYRGNEGTL 355
DB	311 ETALPRRGGVVLGSRHKGVYGR 335
RESULT 11	
T24518	
hypothetical protein T05D4.5 - Caenorhabditis elegans	
C;Species: Caenorhabditis elegans	
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004	
C;Accession: T24518	
R;McMurray, A.	
submitted to the EMBL Data Library, October 1996	
A;Accession: T24518	
A;Reference number: Z19902	
A;Status: preliminary; translated from GB/EMBL/DBJ	
A;Molecule type: DNA	
A;Residues: 1-482 <WIL>	
A;Cross-references: UNIPROT:O45751; UNIPARC:UPI000007C558; EMBL:Z81115; PIDN: CAB03295.1;	
A;Experimental source: clone T05D4	
C;Genetics:	
A;Gene: CESP.T05D4.5	
A;Map position: 3	
A;Introns: 128/3; 184/3; 212/3; 236/3; 305/3	
Query Match	
Best Local Similarity 21.0%; Pred. No. 1.2;	
Matches 78; Conservative 55; Mismatches 146; Indels 93; Gaps 15;	
QY	2 GKSDFLSPK---AJANRIKSGLOKLRYC-----QMCQKOCRDENGFKCH----- 44
DB	118 GKCPYFDQRIITVGRVGEVPVQ--HWFAIYNTSPFPYMFKIKCSNLSFKIHPVCGVL 175
QY	45 -CMSESHQRQLLASENPQPMDFYS-----EPPNDFFLELLRRRFGTKRV 89
DB	176 DALCDASIVTFVGAHPVPPTYSLSIHSLMDPLFLDVQAFRN-----RRNFETRQ- 228
QY	90 HNNIVVNEYISHRHHHNNATQWETLTDFTKWLGREGLCVKDEYTPKGYIQLVIDR----- 144
DB	229 ---FLYVEYAHEDPI-----LVTAEDRDQFHQYMEKVLKTD 264
QY	145 DPETIRRLQLEKKKKQDLD-DEBKTAKFIEQVRRGLEKGEBQETPTVPTETLSRENEBEKV 203
DB	265 EAESFAKKLAEKVERLDKKIDEIKARKIDEKLTIER--EIVRAESPKFEPKTRRLRVEKF 322
QY	204 TPNLNGAGGS-----AGATTSSSSLGPSALKLLGSNASGRK-----KESSQSSAQPAKK 254
A;Residues: 1-375 <BEV>	
A;Cross-references: UNIPROT:Q8SW24; UNIPARC:UPI00000A6F0F; EMBL:AL035523	
A;Experimental source: cultivar Columbia; BAC clone F13M23	
C;Genetics:	
A;Map position: 4	
A;Note: F13M23.160	
Query Match	
Best Local Similarity 23.4%; Pred. No. 0.83;	
Matches 62; Conservative 51; Mismatches 102; Indels 50; Gaps 12;	
QY	132 ETPLKGYIQLVDRDPETIRRLQLEKKKKQDLDEBKTAKFIEQVRRGLEG-----KE 185
DB	80 ETPSG-----DGSNNIYGLNL-RQVENVEKPEPEVKMLKSMRTLESIPDAPEKE 132
QY	186 --QETPV--FTE-----LSRENEEEKVTNFKNGAGSGAGATTSSKSSLG 226
DB	133 DPESFPVDGFEALLAGYWKPGQIGLKAKEDVKIVEKWSGNEGFGFGKSDKAMKMI 192
QY	227 PSALKLLGSNLS-----GKRKSSQSSAQPAKKKKSALD-----EIMELEE---EKKRT 272
DB	193 DN--KLVGSGSHEVKVGINKEIMBKDRVVRKRNRETEGESTEVKAACKQNTVRGQTR 250
QY	273 ARTDAMLQPGIVVYKIITKKL-GEKYHKKKGVVKEVIDRYTAVVVKMTDSGDRKL-LDQTHL 330
DB	251 REKTSWLRSHIKVRIISKDVKGRLYKKAIVTDVVGPTSCDIAMDETQLVQGIQELL 310
QY	331 ETVIPAPGRVVLVNGYRGNEGTL 355
DB	311 ETALPRRGGVVLGSRHKGVYGR 335
RESULT 12	
M2 protein precursor - Streptococcus pyogenes	
C;Species: Streptococcus pyogenes	
A;Variety: serotype M2	
C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004	
C;Accession: S23325; S35761; S61078; S60785	
R;Bessen, D.E.; Fischetti, V.A.	
Infect. Immun. 60, 124-135, 1992	
A;Title: Nucleotide sequences of two adjacent M or M-like protein genes of group A streptococcus	
A;Reference number: S23325; MUID: 92104662; PMID: 1370269	
A;Accession: S23325	
A;Status: preliminary	
A;Molecule type: DNA	
A;Residues: 1-407 <BES>	
A;Cross-references: UNIPROT:P50468; UNIPARC:UPI000012EABE; EMBL:X61276; NID:G47369; PIDN:	
R;Podbielski, A.	
submitted to the EMBL Data Library, November 1992	
A;Reference number: S35761	
A;Accession: S35761	
A;Status: preliminary	
A;Molecule type: DNA	
A;Residues: 1-17 <POD>	
A;Cross-references: UNIPARC:UPI000017823B; EMBL:X69324	
R;Whitmore, A.; Kapur, V.; Sullivan, D.; Musser, J.; Kehoe, M.	
submitted to the EMBL Data Library, July 1994	
A;Description: Noncongruent relationships between variation in emm gene sequences and the pop	
A;Reference number: S60784; MUID: 95198537; PMID: 7891551	
A;Accession: S60785	
A;Status: preliminary; nucleic acid sequence not shown	
A;Molecule type: DNA	
A;Residues: 31-89 <WH2>	
A;Cross-references: UNIPARC:UPI000017823C; EMBL:U11958	
C;Superfamily: M5 protein	
Query Match	
Best Local Similarity 21.0%; Pred. No. 1.3;	
Matches 71; Conservative 69; Mismatches 149; Indels 49; Gaps 11;	
QY	70 EEFNDFLELLRRRFGTKRVHNNIVVNEYISHRE-HIHMNATQWETLTDFTKWLGREGLC 128
DB	90 EEKKDHEKLEKSESDVERHYLQLDQEQYEQQRQKLELELRQSQREVQRY-QEQLQ 148
QY	129 KVDETPKGWYIQVIDRDPETIRRLQLEKKKKQDL-----DEETAKFIEQVTRGL 181
DB	149 KQOLEKEKQISEASR--KSLRRDLEASRAAKKQLEAEHQKLEKEQISEASRSKSLRDL 206
QY	182 E-GKEQETPTVPTETLSRENEBEKVTFNFKNGAGSGAGATTSSSSLGPSALKL-----LGS 235
DB	207 EAGRAAKKQLEAEHQKLEKEQISEASRQGLSRDLASRAAKKQLEAEHQKLEKEQISE 266
QY	236 A-----ASGRKRESSQSAQPAKKKKSALDEI-MELEBEKKTARTDAMLQPGIV 284

Db 267 ASRQGLSRDLASRAKKVVEADLAENSKLQALEKINKELEBEGKKLSKEKAELOAKLE 326
QY 285 VKIITKKLGEKHHKKGVVKEVIDRYTAVVVKMTDGDRLKLDQTHLETVIPAGKRVVLV 344
Db 327 AE--AKALKEQLAKQAEELAKLKGNOTPNAKVAPQANRSRSAMTQKRTLPGTG----- 378
QY 345 NGYRGNEGTLBSINE--KAFSATIVETGPKGRARVE 380
Db 379 -----ETANPFPTAAATVMSAGMLAKRKE 405
RESULT 13
A84671
hypothetical protein At2g27280 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: A84671
R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.B.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84671
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-764 <STO>
A:Cross-references: UNIPROT:Q9XINS; UNIPARC:UPI000017A6B9; GB:AE002093; NID:G5306269; PI
A:Gene: At2g27280
A:Map position: 2
Query Match 6.3%; Score 128.5; DB 2; Length 764;
Best Local Similarity 21.0%; Pred. No. 2.8;
Matches 74; Conservative 69; Mismatches 116; Indels 93; Gaps 16;
QY 46 MESHORQLLASNPQQF-----MDYFSEEPFNDFLELLRRFGTK 87
Db 54 IBEQHKAL---EEDPSAFSYDEVYDDMKQKAVLPRMQD--REERKPRYQNLKMQAER 108
QY 88 RVHNNIVNVEYISHR-----EHTHMATQWET-----LTDFTKWLGRGLCKVDETPKGM 137
Db 109 EKEHEIVRYERKLAKREKDEHLFSOKKEFVTGAYKRLKEEQKKLWAEELRELR----- 163
QY 138 YIQYIDRDDETIRRLQ-----ELEKKKKQLDDDEKTAKEETEEQVRRGL 181
Db 164 -----ERDDVTKKKLDSDFFYFNIGNVAFGAREVEAKEAEKLEEQKAEKLEEQKAEKL 218
QY 182 EKEQETPVFTLSRENEBEKVTFNINLKAGGSAGATTSSKSSSLGP-----SALKLGS 235
Db 219 EELRKEV-TRVEKGRKSPEKV--SPDSGEFGS-----SRKSLSLEPLEAEQAVSEKMG 270
QY 236 AASGRKKESSQSSQAQPAK---KKSALDEIMELEBEKKRTARTDAWLQPGIVVKLIITKK 291
Db 271 DGEERKKSIRKAAKEVPAKINDQKRDEDAIAAKE--RFLARST-----ITNA 318
QY 292 LGEKHHKKGVVKEV--IDRYTAVVKMTDSDGRKL--LDQTHLETVIPAGK 339
Db 319 IGLVKNKEENEQRKIPYVEMHIVRVSINVLDQIEWYICKRFLNVLDFIGK 370
RESULT 14
T34036
hypothetical protein B0041.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T34036
R.Fulton, R.; Wohlmann, P.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid B0041.
A:Reference number: Z21466
A:Accession: T34036

A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1359 <FUL>
A:Cross-references: UNIPROT:Q9U7E0; UNIPARC:UPI0000060B63; EMBL:AF000196; PIDN:AAC24256.1
A:Experimental source: strain Bristol N2; clone B0041
C:Genetics:
A:Gene: CESP:B0041.7
A:Map position: 1
A:Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3
Query Match 6.3%; Score 127; DB 2; Length 1359;
Best Local Similarity 22.3%; Pred. No. 7;
Matches 63; Conservative 47; Mismatches 107; Indels 66; Gaps 9;
QY 145 DPETIRROLELEKK--KKQDLDDDEKTAKEETEEQVRRGLEKEQETPVFTLSRNEBEK 202
Db 214 DEKEVKSKKKSKVVKVKESEDEAPEKKKTEKKRSKTSSESE--SEKSDSEEBEK 271
QY 203 VTFNLKAGGSAGATTSSKSSSLGPSALKLGSASGKRKSSQSSAOPAKKKKSALDEI 262
Db 272 -----ESSPPKPKKKPLAVKKL---SSDESEESDVEVLFPQKKRGAUTLI 314
QY 263 MELEBEKKRTARTDA-----WLOPGIVVKIITKKLGEKHK 299
Db 315 SDSEDEKDKSESEASDVEEKVKKAKKQSESSGSDSSGSIIVNRKSKKKEKPEKKK 374
QY 300 KGWV-----KEVIDRYTAVVKMTDSDGRKLKLDQTHLETVIPAGKRVVLVNGYRGNE 352
Db 375 KGIIMDSKLOKETIDAERA---EKERRKRLKKQKFNFGIVLBERGEDLITMLTG--TSSQ 430
QY 353 GTLESINEKAFSATIVETGP-----LKGRRVVEGIQY 384
Db 431 RKLKSVLDPDSDTVDSESKPVEVHNSLVRLKPKHQAHGIQF 473
RESULT 15
A26655
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Dictyostelium discoideum
C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 09-Jul-2004
C:Accession: A26655; A24728; S00250
R.Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
A:Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium discoi
A:Reference number: A26655; MUID:87092266; PMID:3540939
A:Accession: A26655
A:Molecule type: DNA
A:Residues: 1-2116 <WAR>
A:Cross-references: UNIPROT:P08799; UNIPARC:UPI000012FBD3; GB:M11938; NID:G16;
R.DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
A:Reference number: A24728; MUID:86016788; PMID:3901008
A:Accession: A24728
A:Molecule type: mRNA
A:Residues: 2035-2116
A:Cross-references: UNIPARC:UPI0000177634
R.Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A:Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium n
A:Reference number: S00250; MUID:88112226; PMID:2828113
A:Accession: S00250
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1734-1893 <WAG>
A:Cross-references: UNIPARC:UPI0000177635
C:Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phosph
F.1-818/Domain: globular head <HED>
F.89-747/Domain: myosin motor domain homology <MMOT>
F.179-186/Region: nucleotide-binding motif A (P-loop)
F.819-2116/Domain: alpha-helical rod <ROD>

```
Query Match      6.3%; Score 127; DB 2; Length 2116;
Best Local Similarity 27.0%; Pred. No. 12;
Matches 64; Conservative 37; Mismatches 72; Indels 64; Gaps 11;

QY 129 KVDETPKGWYIQVIDRDPETIRRLQLELE-----KQKKQDLDDDEKTAKEEQV 177
Db 1171 KIKSTLEG-----EVARLQGELEAEQAKSNVEKQKKVELDLEDKSAQLAETA 1220

QY 178 RRGLEK-----EQE-TPVFTLSRENEEEKVTFNLKAGGAGGAGATTSSSSLGPSALK 231
Db 1221 AKQALDKLKKLEQELSEVQTOLSEANNK-----NVNSD-----STNKHLETSENNLK 1268

QY 232 LLGSAASGKRKSSQSSAOPAKKKKSDALDEIM-----ELEEEKRTARTD----- 276
Db 1269 L-----ELEAEQAKAKALEKKRGLGSELEKHVNEQLEEEKQKESNEKRVDEKEV 1320

QY 277 AWLQPGIVVKIITKK-LGEKYHKKGWKEVIDRYTAVVKMTD-SGDRKLKLDQTHLE 331
Db 1321 SELKQIEBEVASKKAVTEAKNKESELEDEIKQYADVSSRDKSVEQLKTLQAKNE 1377
```

Search completed: November 23, 2005, 16:34:28
Job time : 27.9362 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:06:33 ; Search time 130.167 Seconds
(without alignments)
1319.820 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 2028
Sequence: 1 MKSDPLSPKATANRIKSG.....GPKGRVGEIQVEDISKLA 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005as.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2020	99.6	391	2	AAY23619 Murine ki
2	1895	93.4	393	2	AAY23620 A human k
3	1895	93.4	393	7	ADJ69659 Human hea
4	1456	71.8	291	2	AAY23617 Mouse tru
5	1365	67.3	293	2	AAY23618 A human t
6	1245	61.4	390	8	ADP22452 Sea-squir
7	1052.5	51.9	390	4	ABB71416 Drosophil
8	988	48.7	411	3	AG42576 Arabidops
9	913	45.0	424	3	AB03063 Maize KIN
10	913	45.0	437	8	ADX96318 Arabidops
11	913	45.0	437	8	ADX78817 Plant ful
12	901.5	44.5	423	3	AB03064 Maize KIN
13	901.5	44.5	423	3	AB03065 Maize KIN
14	901.5	44.5	423	8	ADT58422 Plant pol
15	843	41.6	382	3	AG42577 Arabidops
16	740	36.5	366	3	AG42578 Arabidops
17	607	29.9	340	3	AG46711 Arabidops
18	607	29.9	343	3	AG46710 Arabidops
19	590	29.1	302	3	AG46712 Arabidops
20	361	17.8	84	4	AG74690 Human col
21	183	9.0	82	3	AG19349 Arabidops
22	170	8.4	81	3	AG19350 Arabidops
23	162	8.0	76	3	AG19351 Arabidops
24	152	7.5	76	3	AG46729 Arabidops

25	149	7.3	278	8	ADX77206	Adx77206 Plant ful
26	142.5	7.0	517	2	AAR22904	Aar22904 1-Caldeem
27	134	6.6	375	3	AAG21445	Aag21445 Arabidops
28	134	6.6	375	8	ADT56172	Adt56172 Plant pol
29	131	6.5	375	3	AAG43126	Aag43126 Arabidops
30	131	6.5	450	3	AAG50080	Aag50080 Arabidops
31	130.5	6.4	1212	6	ABP97029	Abp97029 Rat L-FIL
32	129.5	6.4	1141	4	AAG85008	Aag85008 Shrimp wh
33	128.5	6.3	258	3	AAG43128	Aag43128 Arabidops
34	128.5	6.3	258	3	AAG21447	Aag21447 Arabidops
35	128.5	6.3	263	3	AAG21446	Aag21446 Arabidops
36	128.5	6.3	263	3	AAG43127	Aag43127 Arabidops
37	128.5	6.3	333	3	AAG50082	Aag50082 Arabidops
38	128.5	6.3	338	3	AAG50081	Aag50081 Arabidops
39	128.5	6.3	407	8	ADP49328	Adp49328 S pyrogen
40	125.5	6.2	323	3	AAG15386	Aag15386 Arabidops
41	124.5	6.1	1080	8	ADS23527	AdS23527 Bacterial
42	124	6.1	237	3	AAG15388	Aag15388 Arabidops
43	124	6.1	246	3	AAG15387	Aag15387 Arabidops
44	124	6.1	726	6	ABP96033	Abp96033 Human sur
45	124	6.1	2349	6	AAO16359	Aao16359 Human tra

ALIGNMENTS

RESULT 1
AAY23619
ID AAY23619 standard; protein; 391 AA.
XX
AC AAY23619;
XX
DT 07-SEP-1999 (first entry)
XX
DE Murine kin17 protein.
XX
KW Mouse; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme.
XX
OS Mus sp.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
WPI; 1999-359999/31.
XX
New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.
XX
Claim 19; Page 36-37; 69pp; French.
XX
The present sequence represents a murine kin17 protein. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation
XX
Sequence 391 AA;
SQ

PA (BUCK-) BUCK INST AGE RES.
 XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
 PI Warnock DE;
 XX WPI; 2003-845369/78.
 XX Identifying a mitochondrial target for drug screening assays and for
 PT treating diseases associated with altered mitochondrial function,
 PT comprises detecting a modified polypeptide in a sample and correlating
 PT with the disease.
 XX Claim 1; SEQ ID NO 1465; 180pp; English.
 XX This invention relates to novel mitochondrial targets that can be used
 CC for therapeutic intervention in treating a disease associated with
 CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, nootropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
 CC cytoskeletal activities. This polypeptide sequence is a human heart
 CC mitochondrial protein of the invention.
 XX Sequence 393 AA;
 SQ

Query Match 93.4%; Score 1895; DB 7; Length 393;
 Best Local Similarity 92.4%; Pred. No. 1.8e-159;
 Matches 363; Conservative 14; Mismatches 14; Indels 2; Gaps 1;
 QY 1 MGKSDFLSPKAIANRIKSGKGLQKLRWYCOMCQCRDENGFKCHCMSESHORQLLASSEN 60
 DB 1 MGKSDFLTPKAIANRIKSGKGLQKLRWYCOMCQCRDENGFKCHCMSESHORQLLASSEN 60
 QY 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIMHNATQWETLTDFTK 120
 DB 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIMHNATQWETLTDFTK 120
 QY 121 WLGRGLCKVDETPKGWYIQVIDRDPETIRRQLELEKKKKQDLDDEBKTAKFIEQVRRG 180
 DB 121 WLGRGLCKVDETPKGWYIQVIDRDPETIRRQLELEKKKKQDLDDEBKTAKFIEQVRRG 180
 QY 181 LEGKEQETPVFTELSRNEBEKVTFFNLKNGAGGSAGATTSSKSLGPSALKLLGSAASGK 240
 DB 181 LEGKEQETPVFTELSRNEBEKVTFFNLKNGAGGSAGATTSSKSLGPSALKLLGSAASVK 240
 QY 241 RKSSQSSAOPAKKKKKSALDEIMELEBEKKRTARTDAMIQPGIVVKIITTKLGEKYHK 298
 DB 241 RKSSQSSQTSQSEKKKKKSALDEIMELEBEKKRTARTDAMIQPGIIVKIIITTKLGEKYHK 300
 QY 299 KKGVVKEVIDRYTAVVKNMDSGDLKLDQTHLETIVIPAPGRKRVLVNGYRGNEGTLIESI 358
 DB 301 KKAIVKEVIDRYTAVVKNMDSGDLKLDQTHLETIVIPAPGRKRVLVNGYRGNEGTLIESI 360
 QY 359 NEKAFSATIVETGPKGRRVEGIQYEDISKLA 391
 DB 361 NEKTFSATIVETGPKGRRVEGIQYEDISKLA 393

RESULT 4
 AAY23617
 ID AAY23617 standard; protein; 291 AA.
 XX
 AC AAY23617;
 XX
 DT 07-SEP-1999 (first entry)
 XX

DE Mouse truncated kin17 protein.
 XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme.
 XX Mus sp.
 OS
 XX FR2772046-A1.
 PN
 XX 11-JUN-1999.
 PD
 XX 09-DEC-1997; 97FR-00015536.
 PF
 XX 09-DEC-1997; 97FR-00015536.
 PR
 XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 PA
 XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 PI WPI; 1999-359999/31.
 XX
 DR New DNA coding for human kin17 protein - useful for controlling cell
 PT proliferation or fertility.
 PT
 XX Claim 14; Page 33-34; 69pp; French.
 PS
 XX The present sequence represents a mouse kin17 protein with amino acids
 CC 129-228 deleted. The mammalian kin17 protein is useful for preparing a
 CC medicament for controlling cell proliferation or for controlling
 CC fertility. The medicaments can also be used to treat hyperproliferative
 CC diseases. Fragments between amino acids 55 and 235 (preferably between
 CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
 CC regulating the interaction between proteins and curved DNA. The fragment
 CC can be used to block replication of HIV or its integration into the human
 CC genome or to target repair enzymes to curved DNA sites. Expression
 CC vectors for kin17 can be used for controlling cell proliferation
 XX
 SQ Sequence 291 AA;
 Query Match 71.8%; Score 1456; DB 2; Length 291;
 Best Local Similarity 74.4%; Pred. No. 1.3e-120;
 Matches 291; Conservative 0; Mismatches 0; Indels 100; Gaps 1;
 QY 1 MGKSDFLSPKAIANRIKSGKGLQKLRWYCOMCQCRDENGFKCHCMSESHORQLLASSEN 60
 DB 1 MGKSDFLSPKAIANRIKSGKGLQKLRWYCOMCQCRDENGFKCHCMSESHORQLLASSEN 60
 QY 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIMHNATQWETLTDFTK 120
 DB 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIMHNATQWETLTDFTK 120
 QY 121 WLGRGLCKVDETPKGWYIQVIDRDPETIRRQLELEKKKKQDLDDEBKTAKFIEQVRRG 180
 DB 121 WLGRGLCKVDETPKGWYIQVIDRDPETIRRQLELEKKKKQDLDDEBKTAKFIEQVRRG 180
 QY 181 LEGKEQETPVFTELSRNEBEKVTFFNLKNGAGGSAGATTSSKSLGPSALKLLGSAASGK 240
 DB 181 LEGKEQETPVFTELSRNEBEKVTFFNLKNGAGGSAGATTSSKSLGPSALKLLGSAASGK 240
 QY 241 RKSSQSSAOPAKKKKKSALDEIMELEBEKKRTARTDAMIQPGIVVKIITTKLGEKYHK 300
 DB 241 RKSSQSSAOPAKKKKKSALDEIMELEBEKKRTARTDAMIQPGIIVKIIITTKLGEKYHK 300
 QY 301 GVVKEVIDRYTAVVKNMDSGDLKLDQTHLETIVIPAPGRKRVLVNGYRGNEGTLIESI 360
 DB 201 GVVKEVIDRYTAVVKNMDSGDLKLDQTHLETIVIPAPGRKRVLVNGYRGNEGTLIESI 260
 QY 361 KAFSATIVETGPKGRRVEGIQYEDISKLA 391
 DB 261 KAFSATIVETGPKGRRVEGIQYEDISKLA 291

RESULT 5
AAY23618
ID AAY23618 standard; protein; 293 AA.
XX
XX AAY23618;
AC
XX
DT 07-SEP-1999 (first entry)
XX
DE A huamn truncated kin17 protein.
XX
XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW protein interaction; curved DNA; HIV replication; HIV integration;
KW repair enzyme.
XX
OS Homo sapiens.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
XX 09-DEC-1997; 97FR-00015536.
PF
XX 09-DEC-1997; 97FR-00015536.
PR
XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
PA
XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
PI WPI; 1999-359999/31.
XX
XX New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.
XX
XX Claim 15; Page 34-35; 69pp; French.
PS
CC The present sequence represents a truncated human kin17 protein with
CC amino acids 128-228 deleted. The mammalian kin17 protein is useful for
CC preparing a medicament for controlling cell proliferation or for
CC controlling fertility. The medicaments can also be used to treat
CC hyperproliferative diseases. Fragments between amino acids 55 and 235
CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein
CC are useful for regulating the interaction between proteins and curved
CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
CC proliferation
XX
SQ Sequence 293 AA;

Query Match 67.3%; Score 1365; DB 2; Length 293;
Best Local Similarity 69.2%; Pred. No. 1.6e-112;
Matches 272; Conservative 9; Mismatches 10; Indels 102; Gaps 2;
Qy 1 MGKSDFLSPKATANRIKSKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQRLILASEN 60
Db 1 MGKSDFLTPKATANRIKSKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQRLILASEN 60
Qy 61 PQOFMDYFSEEFNRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATQWETLTDFTK 120
Db 61 PQOFMDYFSEEFNRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATQWETLTDFTK 120
Qy 121 WLGREGLCKVDETPKGWYIQYIDRDPETIRRQLEBKKKQDLDDBEKTAKEEQVRG 180
Db 121 WLGREGLC----- 128
Qy 181 LSGKEQETPVFTELSRNEEEKVTFNLNKGAGSGAGATTSSKSSLGPSALKLLGSAAGK 240
Db 129 -----ALKTIGSASVK 140
Qy 241 RKSSQSSAQ--PAKKKKSALDEIMEBEKKKRTARTDQWPGIVVVKIITTKLGEKYHK 298
Db 141 RKSSQSSQTSQSKKKKKKKSALDEIMEBEKKKRTARTDQWPGIIVVVKIITTKLGEKYHK 200

Qy 299 KGVVKEVIDRYTAVVVMKTTDSGDRLLKLDQTHLETVIPAGKRVLVNGGYRGNEGTFESI 358
Db 201 KKAIVKEVIDRYTAVVVMKIDSGDKLKDQTHLETVIPAGKRILVNGGYRGNEGTFESI 260
Qy 359 NEKAFSATIVTETGPKGRRVEGIIQYEDISKLA 391
Db 261 NEKTFATIVTETGPKGRRVEGIIQYEDISKLA 293
RESULT 6
ADP22452
ID ADP22452 standard; protein; 390 AA.
XX ADP22452;
XX
DT 12-AUG-2004 (first entry)
XX
DE Sea-squirt (Ciona intestinalis) zinc finger protein #13.
XX
KW sea-squirt; zinc finger protein; gene detection; drug development;
KW zinc finger protein-associated disease.
XX
OS Ciona intestinalis.
XX
PN JP2004057126-A.
XX
PD 26-FEB-2004.
XX
PF 31-JUL-2002; 2002JP-00222484.
XX
PR 31-JUL-2002; 2002JP-00222484.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
DR WPI; 2004-208711/20.
DR N-PSDB; ADP22451.
XX
PT Novel gene encoding zinc finger protein, useful as probe in gene
PT detecting instruments and in development of drug for treating zinc finger
PT protein associated diseases.
XX
PS Claim 1; SEQ ID NO 26; 972pp; Japanese.
XX
CC The invention comprises the amino acid and coding sequences of sea-squirt
CC (Ciona intestinalis) zinc finger proteins. The DNA and protein sequences
CC of the invention are useful in a gene detecting instrument. The DNA and
CC protein sequences of the invention are useful in the development of drugs
CC for the treatment of zinc finger protein-associated diseases. The present
CC amino acid sequence represents a sea-squirt zinc finger protein of the
CC invention.
XX
SQ Sequence 390 AA;

Query Match 61.4%; Score 1245; DB 8; Length 390;
Best Local Similarity 62.8%; Pred. No. 1.1e-101;
Matches 247; Conservative 50; Mismatches 84; Indels 12; Gaps 6;
Qy 1 MGKS--DFLSPKAIANRIKSKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQRLILAS 58
Db 1 MGKEKPGFLTAKAIGNRIKSKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQRLIAG 60
Qy 59 ENPQOFMDYFSEEFNRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATQWETLTD 118
Db 61 ENPGAFHSPFSSFFDFMQLLTKTFGKRVHNNIVYNEYISHREHIHMNATQWTLTAF 120
Qy 119 TKWLREGGLCKVDETPKGWYIQYIDRDPETIRRQLEBKKKQDLDDBEKTAKEEQVR 178
Db 121 TKWLREGHCKVDQTEKGFQYIDRDLVLARQELQKQREDDDERQAKAIEEMVK 180
Qy 179 RGLG-GRKQETPVFTELSRNEEEKVTFNLNKGAGSGAGATTSSKSSLGPSALKLLGSA 237
Db 181 RQDQTAQSDSGKFTLVLR-NDEEKIATNLQWNSPKA-----ATSSQLAP---RVSLSL 232

PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.

PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149428P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149830P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.

```

PR 26-OCT-1999; 99US-01611360P.
PR 26-OCT-1999; 99US-01611361P.
PR 28-OCT-1999; 99US-01611920P.
PR 28-OCT-1999; 99US-01611922P.
PR 28-OCT-1999; 99US-01611932P.
PR 28-OCT-1999; 99US-01611933P.
PR 28-OCT-1999; 99US-0162142P.

Query Match 48.7%; Score 988; DB 3; Length 411;
Best Local Similarity 48.8%; Pred. No. 8.7e-79;
Matches 203; Conservative 67; Mismatches 116; Indels 30; Gaps 7;

Qy 1 MGKSDFLSPKAIANRIKSKGLQKLRWYQCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
1 MGKNDFLTPKAIANRIKAKGLQKLRWYQCMQCKQCRDENGFKCHCMSESHORQMVFGQN 60

Qy 61 PQQFMDYFSEFRNDFLELLRRFGTKRVNNIVYNEYISHREHHNATQWETLTDFTK 120
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
61 PTRVVDGYSEFEQTFLDLRRSHRFSRIAATVVYNEYINDRHHVHMNSTEWATLTFEIK 120

Qy 121 WLREGGLCKVDETPKGWYIOYIDRDPETIRROLELEKKKQDLDDEKTAKEEQQVRRG 180
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
121 HLKGTGCKVEETPKGWFYIYIDRSETLFRKLRKNRVKSDLAEBEKQREBQRIQIERA 180

Qy 181 LE-----GKQETPVFTLSRENEE-----KVTFLNKGAGGSAGATTSKS 222
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
181 AEKLNCGGGEGETSGNDEVDDGDERKDBDLRLKSGVKVGFAL---GGGVKQVATGKE 237

Qy 223 SSLGPSALKULGSAAS-----GKRKSSQSAQPAKKKKKALDEIMELEBEKK-RTARTD 276
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
238 R--GESSKLLFGDBENDKVERGEKRRKSGDGRSEKERRSALDELMKEBEKKERMNRKD 295

Qy 277 AWLQPGIVVKIITKLGKEK-YHKKGVVKEVIDRYTAVVWMTSGDRLKLDQTHLETVIP 335
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
296 YWLFEGHIVKMSKALAEKGYIKQGVVKKVIDNYVGEIKMLDSKHVLRVDQKELETVLP 355

Qy 336 APGRKRVLVNGYRGNEGTLSEINEKAFSATVIETGPKLGRRVEGIQYEDISKLA 391
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
356 QIGGVKIVNGAYRGSNARLLGVDTKEFCAKVQIEKGYVIGRVIKSIEYEDICKLA 411

RESULT 9
AAB03063
ID AAB03063 standard; protein; 424 AA.
XX AC AAB03063;
XX DT 27-SEP-2000 (first entry)
XX DE Maize KINI17 orthologue, ZmKINH-1.
XX KW ZmKINH-1; KINI17 orthologue; maize; zinc finger protein; RecA homologue;
KW nuclear localisation; nonhomologous recombination;
KW illegitimate recombination; double stranded DNA binding; curved DNA;
KW homologous gene targeting; transgenic plant.
XX OS Zea mays.
XX FN WO200024900-A1.
XX PD 04-MAY-2000.
XX PF 06-OCT-1999; 99WO-US023280.
XX PR 27-OCT-1998; 98US-0105802P.
XX PA (PION-) PIONEER HI-BRED INT INC.
XX PI Mahajan PB;
XX DR WPI; 2000-350741/30.
XX AC N-PSDB; AAA52589.
XX PT Nucleic acids encoding maize KINI17 orthologue proteins useful for

```

```

PT preventing illegitimate recombination in cells.
XX Claim 11; Page 64-66; 84pp; English.
XX This sequence represents the maize KINI17 orthologue ZmKINH-1. The
CC invention relates to maize KINI17 orthologues (AAB03063-B03065), nucleic
CC acids encoding them (AAA52589-A52591), and expression vectors, transgenic
CC plants and plant seeds comprising nucleotides encoding maize KINI17
CC orthologues. KINI17 has, until now, been found only in animal (avian,
CC rodent and human) cells, this invention being the first report describing
CC the presence of KINI17 in plants. Murine KINI17 was found to have
CC significant homology to Escherichia coli RecA protein, and contains a
CC zinc finger motif and a nuclear localisation signal. KINI17 binds double-
CC stranded DNA, preferentially binding to curved DNA, and forms
CC intranuclear foci on overexpression in mammalian cells. It is also
CC indicated that KINI17 plays a role in non-homologous (illegitimate)
CC recombination, which occurs at higher rates among higher eukaryotes,
CC particularly plants. Illegitimate recombination in plants is a major
CC impediment to the generation of transgenic crops such as maize. Maize
CC KINI17 orthologue nucleic acid sequences may be used to generate
CC transgenic plants. The transgenic plants generated can be monocots or
CC dicots and are particularly maize, soybean, sunflower, sorghum, canola,
CC wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
CC sequences may be used to reduce KINI17 levels in embryogenic callus or
CC embryo cells, thereby reducing the amount of non-homologous recombination
CC and enhancing homologous gene targeting
XX Sequence 424 AA;
SQ

Query Match 45.0%; Score 913; DB 3; Length 424;
Best Local Similarity 44.9%; Pred. No. 4.3e-72;
Matches 190; Conservative 75; Mismatches 126; Indels 32; Gaps 7;

Qy 1 MGKSDFLSPKAIANRIKSKGLQKLRWYQCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
1 MGKHEFLTPKAIANRIKAKGLQKLRWYQCMQCKQCRDENGFKCHCMSESHORQMVFGMA 60

Qy 61 PQQFMDYFSEFRNDFLELLRRFGTKRVNNIVYNEYISHREHHNATQWETLTDFTK 120
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
61 PDRVVEGSEEFLESFLSIRRAHRSRVAA TVYNEYIADRHVHVNSTRWATLTFFVK 120

Qy 121 WLREGGLCKVDETPKGWYIOYIDRDPETIRROLELEKKKQDLDDEKTAKEEQQVRR- 179
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
121 FLGREGYCKVEDTPKGWFYIYIDRSEQAVKRLKRIKSDMADDERQRMIAQIERA 180

Qy 180 -----GLEKQETPVFTLSRENEEKTFFNLNKGAGGSAG--ATTSSSLGSP 227
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
181 HKSLAKPNGGGAAGEPESEGESYSGSDDEQPEDDSKADKATKIALQKAVPGP 240

Qy 228 SA-----LKL-----LGSASGKRK-----SSQSSAQPAKKKKKALDEIMELEBE-K 269
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
241 KVPFDDPKPKFGFDEEDSGTRDQKNELTKMGKDVKAARAKRSALDELMEKEEKAK 300

Qy 270 KRTARTDAWLQPGIVVKIITKLGKEK-YHKKGVVKEVIDRYTAVVWMTSGDRLKLDQDT 328
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
301 ERSNRKDYWLCPGIVVKVMSKLAEGYIKQGVVKKVIDNYVGEIEMLESKHVLRVDQD 360

Qy 329 HLETVIPAPGRVVLVNGYRGNEGTLSEINEKAFSATVIETGPKLGRRVEGIQYEDIS 388
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
361 ELETVIPQIGLVRIVNGAYRGSNARLLSVDTKEFCAKVQIEKGYVIGRVIKSIEYEDIC 420

Qy 389 KLA 391
Db |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
421 KIS 423

RESULT 10
ADX96318
ID ADX96318 standard; protein; 437 AA.
XX AC ADX96318;
XX

```



```
PS Claim 1; SEQ ID NO 48183; 15pp; English.
XX
CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This is the amino acid sequence of a plant full length insert
CC polypeptide that can be used in the recombinant DNA construct of the
CC invention.
XX
SQ Sequence 437 AA;
Query Match 45.0%; Score 913; DB 8; Length 437;
Best Local Similarity 44.9%; Pred. No. 4.5e-72;
Matches 190; Conservative 75; Mismatches 126; Indels 32; Gaps 7;
Qy 1 MGKSDFLSPKAIANRIKSKGLQKLRYCQMCQKCRDENGPKCHCMSESHORQLLASEN 60
Db 14 MGKHEFLTPKAIANRIKAKGLQKLRYCQMCQKCRDENGPKCHCMSESHORQMVFGMA 73
Qy 61 PQQMDYFSEFRNDFLELLRRRFGTKRVHNNIYVYNEIISHREIHNNATOWETLTDFTK 120
Db 74 PDRVVEGFSEEFLESLIRRAHRHGRVAATVYVYNEIADRHVHNNSTRTWATLTDFVK 133
Qy 121 WLREGGLCKVDTPKGWYIOVDRPETIRROLELEKKKQDLDDEETAKFIEQVRR- 179
Db 134 FLREGYCKVEDTPKGWYIYDRDSEQAVKRLKRIKSKDMADDERQERMIARQIERA 193
Qy 180 -----GLEGEQETPVFTLSRENEEEKVTNFKGAGGSAG--ATTSSSSSLGP 227
Db 194 HKSLAKPNGGAAEGEPESGSEYSGSDDEQBPEDDSKEADKATGKIAIALQKAVRGP 253
Qy 228 SA-----IKL-----LGSAAGSKRK-----SSQSAQPAKKKKSALDIMELEE-K 269
Db 254 KVNPFDDPKMKFGFDEEDSGTRDQEKNEVTKMGKDKVKAAGKRSALDELMEEEKAK 313
Qy 270 KRTARTDALQPGIWKIITTKLGBK-YHKKGVVYKEVIDRYTAVVKMTDSGDRLLKLDQT 328
Db 314 ERSNRKDYWLCPGIIVKVNKSLAEKGYKQGVVKKVIDKYVGEIEMLESHVLRVDQD 373
Qy 329 HLETVIPAGKRVLVNGYRGNEGTLSINEKAFSATIVTETPLKGRVVEGIQYEDIS 388
Db 374 ELETVPIQGLVIRVNGAYRGSNARLLSVDTEKFCAKVQVEKGLYDGKVLRAVEYEDIC 433
Qy 389 KLA 391
Db 434 KIS 436
RESULT 12
AAB03064
ID AAB03064 standard; protein; 423 AA.
XX
AC AAB03064;
XX
DT 27-SEP-2000 (first entry)
XX
DE Maize KIN17 orthologue, ZmKINH-2.
XX
KW ZmKINH-2; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
KW nuclear localisation; nonhomologous recombination;
KW illegitimate recombination; double stranded DNA binding; curved DNA;
KW
homologous gene targeting; transgenic plant.
Zea mays.
WO2000024900-A1.
04-MAY-2000.
06-OCT-1999; 99WO-US023280.
27-OCT-1998; 98US-0105802P.
(PION-) PIONEER HI-BRED INT INC.
Mahajan PB;
WPI; 2000-350741/30.
N-PSDB; AAS2590.
Nucleic acids encoding maize KIN17 orthologue proteins useful for
preventing illegitimate recombination in cells.
Claim 11; Page 64-66; 84pp; English.
This sequence represents the maize KIN17 orthologue ZmKINH-2. The
invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic
acids encoding them (AAS2589-A52591), and expression vectors, transgenic
plants and plant seeds comprising nucleotides encoding maize KIN17
orthologues. KIN17 has, until now, been found only in animal (avian,
rodent and human) cells, this invention being the first report describing
the presence of KIN17 in plants. Murine KIN17 was found to have a
significant homology to Escherichia coli RecA protein, and contains a
zinc finger motif and a nuclear localisation signal. KIN17 binds double-
stranded DNA, preferentially binding to curved DNA, and forms
intracellular foci on overexpression in mammalian cells. It is also
induced on exposure to gamma or ultraviolet radiation. These findings
indicate that KIN17 plays a role in non-homologous (illegitimate)
recombination, which occurs at higher rates among higher eukaryotes,
particularly plants. Illegitimate recombination in plants is a major
impediment to the generation of transgenic crops such as maize. Maize
KIN17 orthologue nucleic acid sequences may be used to generate
transgenic plants. The transgenic plants generated can be monocots or
dicots and are particularly maize, soybean, sunflower, sorghum, canola,
wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
sequences may be used to reduce KIN17 levels in embryogenic callus or
embryo cells, thereby reducing the amount of non-homologous recombination
and enhancing homologous gene targeting
Sequence 423 AA;
Query Match 44.5%; Score 901.5; DB 3; Length 423;
Best Local Similarity 44.5%; Pred. No. 4.5e-71;
Matches 189; Conservative 72; Mismatches 127; Indels 37; Gaps 8;
Qy 1 MGKSDFLSPKAIANRIKSKGLQKLRYCQMCQKCRDENGPKCHCMSESHORQLLASEN 60
Db 1 MGKHEFLTPKAIANRIKAKGLQKLRYCQMCQKCRDENGPKCHCMSESHORQMVFGMA 60
Qy 61 PQQMDYFSEFRNDFLELLRRRFGTKRVHNNIYVYNEIISHREIHNNATOWETLTDFTK 120
Db 61 PDRVVEGFSEEFLESLIRRAHRHGRVAATVYVYNEIADRHVHNNSTRTWATLTDFVK 120
Qy 121 WLREGGLCKVDTPKGWYIOVDRPETIRROLELEKKKQDLDDEETAKFIEQVRRG 180
Db 121 LLREGYCKVEDTPKGWYIYDRDSEQAVKRLKRIKSKDMADDERQERMIARQIERA 180
Qy 181 LE-----GKEQETPVFTLSRENEE-EKVTNFKGAGGSAG--TTS 220
Db 181 HKSLAKPNGGAAEGEPESGSEYSGSDDEQBPEDDSKEADKATGKIAIALQKAVR 237
Qy 221 KSSSLGSPALK-----LGSAAGSKRK-----SSQSAQPAKKKKSALDIMELEE- 267
Db 238 RGPKNPPEPKVFRFDEEDGLTGTEKDELAKKKGDKVKAADARRSALDELMEKEEM 297
```

Qy	268	EKKTARTDAWLQPGIVVKIITKYLGEK-YHKKKGVWKEVIDRYTAVVKMTDSGDRLLKLD	326
Db	298	AKERSNRKDYWLCPGIVVKVMSKSLAEKGYRKQGVVRKVMDDKYVGEIEMLESKHVLRVD	357
Qy	327	QTHLETVIPAPQKRVLVNGGVRNGEGTLESINEKAFSATIVIEGTPLKGRVVEGIIQYED	386
Db	358	QDELETVIPQIGGLVRIILNGAYRGSNARLLSVDTKFCAKVQVEKGLYDGKVLRAVKYED	417
Qy	387	ISKLA 391	
Db	418	ICKIS 422	
RESULT 13			
AAB03065			
ID	AAB03065	standard; protein; 423 AA.	
XX	XX	AAB03065;	
XX	XX		
DT	27-SEP-2000	(first entry)	
XX	XX		
DE	Maize	KIN17 orthologue, ZmKINH-3.	
XX	XX		
KW	ZmKINH-3;	KIN17 orthologue; maize; zinc finger protein; RecA homologue;	
KW	nuclear localisation;	nonhomologous recombination;	
KW	illegitimate recombination;	double stranded DNA binding; curved DNA;	
KW	homologous gene targeting;	transgenic plant.	
XX	XX		
OS	Zea mays.		
XX	XX		
PN	WO200024900-A1.		
XX	XX		
PD	04-MAY-2000.		
XX	XX		
PF	06-OCT-1999;	99WO-US023280.	
XX	XX		
PR	27-OCT-1998;	98US-0105802P.	
XX	XX		
PA	(PION-) PIONEER HI-BRED INT INC.		
XX	XX		
FI	Mahajan PB;		
XX	XX		
DR	WPI; 2000-350741/30.		
DR	N-PSDB; AAA52591.		
XX	XX		
PT	Nucleic acids encoding maize KIN17 orthologue proteins useful for		
PT	preventing illegitimate recombination in cells.		
XX	XX		
PS	Claim 11; Page 64-66; 84pp; English.		

CC	and enhancing homologous gene targeting		
XX			
QQ	Sequence 423 AA;		
	Query Match	44.5%;	Score 901.5; DB 3; Length 423;
	Best Local Similarity	44.5%;	Pred. No. 4.5e-71;
	Matches 189; Conservative	77;	Mismatches 117; Indels 37; Gaps 8
QY	1	MGKSDFLSPKAIANRIKSGLOKLRLWYQMCQKQCRDENGPKCHCMSES	SHORQQLLAS
DB	1	MGKHEFLTPKAIANRIKAKGLQKLRLWYQMCQKQCRDENGPKCHCMSES	SHORQMQVFGMA
QY	61	POQPMDFSEFRNDFLELLRRPGTKRVHNNIVYNEYISHREHIHNNATOWELT	DTFK 120
DB	61	PDRVVEGFSSEFLSEFLSLIRRAHRHRSRVATVVYNEYIADRHVHMNMSTRWATL	TEFVK 120
QY	121	WLREGGLCKVDENPKGWYIOYIDRDPETIRRQLELEKKKKQDLDEEKTAKFIEBQVR	RG 180
DB	121	LLGREGYCKVEDTPKGMFTYIIDRDSQAQVDRLLKRRKIKSDMAEDEKQERMIARQ	IERA 180
QY	181	LE-----GKEQETPVFTELSRENEE-EKVTFLNFKGAGGSAGA--TTS	220
DB	181	HKSLAKPNDGAAEGEPESGEE--YSGSDDDGELEDDVSKKANKATGKIAIALQ	TAV 237
QY	221	KSSSLGPSALK-----LLGSAASKRKE-----SSQSSAQPAKKKKKSA	DEINELBE- 267
DB	238	RGKVPNPFDPKPEVKPRFDEGLGTGEKDELAKKKQKVAADARRSALDEL	MKEBEM 297
QY	268	EKRTARTDNLQGIWVKLIITKKLGK-YHKKKGVVKEVIDRYTAVVM	TDSDGLKLD 326
DB	298	AKERSNRKDWLCPGIVVVMVMSLAEBKGYKQGVVRKVMKDYVGEIEMLES	KHVLKVD 357
QY	327	QTHLETVIPAPGKRVLVNGYRGNEGTLESINEKAFSATIVITG	PLKGRRVEGIOYED 386
DB	358	QDELETVPIQIGGLVRLINGAYRGSNARLLSVDTKEFCAKQV	KEGLYDGKVLRAVKYED 417
QY	387	ISKLA 391	
DB	418	ICKIS 422	
RESULT 14			
ADTS8422	ID	ID	ADTS8422 standard; protein; 423 AA.
XX	XX	XX	
AC	ADTS8422;		
XX	13-JAN-2005 (first entry)		
DT	Plant polypeptide, SEQ ID 8499.		
DE	Plant; transgenic; cold tolerance; growth rate; drought tolerance;		
XX	disease resistance; galactomannan production; plant growth regulator;		
KW	heat tolerance; herbicide tolerance; lignin production;		
KW	extreme osmotic condition tolerance; pathogens resistance;		
KW	pest resistance; yield improvement; seed oil yield; seed protein yield.		
XX	Viridiplantae.		
OS	US2004216190-A1.		
XX	28-OCT-2004.		
PN	18-DEC-2003; 2003US-00739930.		
PD	28-APR-2003; 2003US-00424599.		
PF	28-APR-2003; 2003US-00425115.		
XX	(KOVA/) KOVALIC D K.		
XX	Kovalic DK;		
XX	WPI: 2004-757369/74.		

PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 08-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 41.6%; Score 843; DB 3; Length 382;
Best Local Similarity 45.7%; Pred. No. 6.2e-66;
Matches 177; Conservative 64; Mismatches 116; Indels 30; Gaps 7;

Db	1	MCQKQCEDENGFKCHCNSHQQMVFQGNPTRVVDGYSEEFQTFDLMRSHRFSRI	60
Qy	90	HNNIVYNEYISHREHIHMNATOWETLTDFTYKWLGREGLCKVDETPKGWYIOYIDRDPETI	149
Db	61	AATVVYNEYINDRRHHVHMNSTEWATLTFETIKHLGKTGCKVEETPKGWFITYIDRSETL	120
Qy	150	RRQLELEKKKKQDLDDDEKTAKEEBQVRGLE-----GKEQETPVFTELSRNEEBE---	201
Db	121	FKERLKNYKVSDLAEBEKKQEREIQRIERAALKNGGGGEGGETSGNDEVDDGDDERKK	180
Qy	202	-----KVTFLNKGAGGSAGATTSSKSSILGPSALKLLGSAAS-----GKRRESSQ	246
Db	181	DEDLRLKSGVKVGFAL---GGYKQVATGKER--GESSKLLFGDEENDKVERGEKKRSG	235
Qy	247	SSAQPAKKKXSALDEINELBEEKK-RTARTDAMLQPGIIVVKIITYKLGK- YHKKKGWVK	304
Db	236	DSGRSEKERRSALDELMKEBEKKERNRNDYMLFEGIIVKVMSKALAEKGYKQKGVVK	295
Qy	305	EVIDRYTAVVKMTDSGDRLXLDOTHLETVIPAGKRVLVNLGGYRGNEGTLSEINEKAFS	364
Db	296	KVIDNYVEIKRMLDSKHVLRVDQKELETVLFPQIGGMVKIVNGAYRGSNARLLGVDTBKFC	355
Qy	365	ATIVIETGPLKGRVVEGIQYEDISKLA	391
Db	356	AKVQIEKGVYDGRVKSIEYEDICKLA	382

Search completed: November 23, 2005, 16:28:20
Job time : 134.167 secs

This Page Blank (usp1c)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:19:43 ; Search time 152.61 Seconds
(without alignments)
1807.628 Million cell updates/sec

Title: US-09-555-529-25

Perfect score: 2028

Sequence: 1 MGKSDPLSPKAIANRIKSG.....GPKLGRVRGIIQVEDISKLA 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2028	100.0	391	Q8K339_MOUSE	Q8K339 mus musculus
2	1895	93.4	393	Q60870_HUMAN	Q60870 homo sapien
3	1678	82.7	389	Q6GL97_XENTR	Q6GL97 xenopus tro
4	1646	81.2	387	Q68F56_XENLA	Q68F56 xenopus lae
5	1552	76.5	383	Q5RH25_BRARE	Q5RH25 brachydanio
6	1112	54.8	387	Q7PGA1_ANOGA	Q7PGA1 anopheles g
7	1056	52.1	195	Q9CV58_MOUSE	Q9CV58 mus musculus
8	1052.5	51.9	390	Q9VPH4_DROME	Q9VPH4 drosophila
9	1050.5	51.8	390	Q8SKR2_DROME	Q8SKR2 drosophila
10	988	48.7	411	Q9ZVU5_ARATH	Q9ZVU5 arabidopsis
11	969.5	47.8	404	Q9XWF2_CABEL	Q9XWF2 caenorhabdi
12	932.5	46.0	269	Q4TDV9_TETNG	Q4TDV9 tetraodon n
13	913	45.0	430	Q75LIJ_ORYSA	Q75LIJ oryza sativ
14	839	41.4	402	Q9A187_ORYSA	Q9A187 oryza sativ
15	835	41.2	397	Q60L39_CAEBR	Q60L39 caenorhabdi
16	831.5	41.0	427	Q4S5G2_TETNG	Q4S5G2 tetraodon n
17	825	40.7	445	Q55D16_DICDI	Q55D16 dictyosteli
18	689.5	34.0	244	Q76926_DROME	Q76926 drosophila
19	649	32.0	178	Q6X190_DROYA	Q6X190 drosophila
20	648	32.0	378	Q4N9K7_THEPA	Q4N9K7 theileria p
21	645	31.8	441	Q4XDV6_PLACH	Q4XDV6 plasmodium
22	644	31.8	412	Q4UG63_THEAN	Q4UG63 theileria a
23	644	31.8	442	Q8IKG2_PLAF7	Q8IKG2 plasmodium
24	644	31.8	445	Q4Z7D7_PLABE	Q4Z7D7 plasmodium
25	630	31.1	457	Q7RD49_PLAYO	Q7RD49 plasmodium
26	604	29.8	334	Q4WF60_ASPFU	Q4WF60 aspergillus
27	604	29.8	337	Q4I5W6_GIBZE	Q4I5W6 gibberella
28	590	29.1	335	Q5AQ50_EMENI	Q5AQ50 aspergillus
29	578.5	28.5	304	Q9Y7X9_SCHPO	Q9Y7X9 schizosacch
30	544	26.8	324	Q55M94_CRYNE	Q55M94 cryptococcu
31	544	26.8	324	Q5K8H7_CRYNE	Q5K8H7 cryptococcu

32	533.5	26.3	328	2	Q525W2_MAGGR	Q525w2 magnaporthe
33	520	25.6	358	2	Q4PG11_USTMA	Q4pg11 ustilago ma
34	504.5	24.9	256	2	Q6C2K1_YARLI	Q6c2k1 yarrowia li
35	457.5	22.6	276	2	Q7S742_NEUCR	Q7s742 neurospora
36	429.5	21.2	263	2	Q5CFM3_CRYHO	Q5cfm3 cryptospori
37	429	21.2	265	2	Q5CYD1_CRYPV	Q5cyd1 cryptospori
38	399.5	19.7	238	2	Q51F62_ENTHI	Q51f62 entamoeba h
39	397.5	19.6	271	2	Q6BQO0_DEBHA	Q6bqo0 debaryomyce
40	366	18.0	239	2	Q6CXJ7_KLULA	Q6cxj7 kluyveromyc
41	365	18.0	236	2	Q759P8_ASHGO	Q759p8 ashbya goss
42	361	17.8	121	2	Q4XND8_PLACH	Q4xnd8 plasmodium
43	357.5	17.6	239	2	Q5A673_CANAL	Q5a673 candida alb
44	355	17.5	241	2	Q6FMS8_CANGA	Q6fms8 candida gla
45	304.5	15.0	88	2	Q4TDW0_TETNG	Q4tdw0 tetraodon n

ALIGNMENTS

RESULT 1

Q8K339_MOUSE PRELIMINARY; PRT; 391 AA.

AC Q8K339;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Antigenic determinant of rec-A protein.
GN Name:Kin;

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Smallus D.E.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Marra M.A.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

NUCLEOTIDE SEQUENCE.

RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

[3]

NUCLEOTIDE SEQUENCE.

RP MEDLINE=92020193; PubMed=1923796;
RX Angulo J., Rouer E., Mazin A., Mattei M., Tissier A., Horellou P.,
RA Benarous R., Devoret R.;
RT "Identification and expression of the cDNA of KIN17, a zinc-finger
gene located on mouse chromosome 2, encoding a new DNA binding
protein.";
RL Nucleic Acids Res. 19:5117-5123(1991).
DR EMBL; BC028860; AAH28860.1; -; mRNA.
DR PIR; S18666; S18666.

```
DR Ensembl; ENSMUSG00000037262; Mus musculus.
DR MGI; 96676; Kin.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR006646; KOW_sub.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00467; KOW; 1.
DR SMART; SM00739; KOW; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ SEQUENCE 391 AA; 44722 MW; 955BC6A4FF4D3B6E CRC64;

Query Match 100.0%; Score 2028; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 6.3e-113;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKSDFLSPKAIANRIKSKGLQKLRYWCQKQCRDENGFKCHCMSESHQRLLLASEN 60
DB 1 MGKSDFLSPKAIANRIKSKGLQKLRYWCQKQCRDENGFKCHCMSESHQRLLLASEN 60
QY 61 PQQFMDYFSEEPNDFLELLRRRFGTKRVHNNIVYNEIYSHREHIMHNATQWETLTDFTK 120
DB 61 PQQFMDYFSEEPNDFLELLRRRFGTKRVHNNIVYNEIYSHREHIMHNATQWETLTDFTK 120
QY 121 WLGRGLCKVDTPKGYIYQIDRDPETIRROLELEKKKKQDLDDDEKTAFTIEEQVRG 180
DB 121 WLGRGLCKVDTPKGYIYQIDRDPETIRROLELEKKKKQDLDDDEKTAFTIEEQVRG 180
QY 121 WLGRGLCKVDTPKGYIYQIDRDPETIRROLELEKKKKQDLDDDEKTAFTIEEQVRG 180
DB 121 WLGRGLCKVDTPKGYIYQIDRDPETIRROLELEKKKKQDLDDDEKTAFTIEEQVRG 180
QY 181 LEGKEQETVPVFTSLRENEEEKVTFLNKGAGSAGATTSSKSSLGPSALKLLGSAASGK 240
DB 181 LEGKEQETVPVFTSLRENEEEKVTFLNKGAGSAGATTSSKSSLGPSALKLLGSAASGK 240
QY 241 RKSSQSSQAQ--PAKKKKSALDEIMELBEEKKRTARTDAWLQPGIVVKIITKLGKGYHKK 300
DB 241 RKSSQSSQAQ--PAKKKKSALDEIMELBEEKKRTARTDAWLQPGIVVKIITKLGKGYHKK 300
QY 301 GVVKEVIDRYTAVVKMTSDGRLKLDQTHLETIVIPAGKRVVLNGGYRGNEGTLESINE 360
DB 301 GVVKEVIDRYTAVVKMTSDGRLKLDQTHLETIVIPAGKRVVLNGGYRGNEGTLESINE 360
QY 361 KAFSATIVETGPKRRVEGVIQYEDISKLA 391
DB 361 KAFSATIVETGPKRRVEGVIQYEDISKLA 391

RESULT 2
O60870 HUMAN PRELIMINARY; PRT; 393 AA.
AC O60870;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Kin17 protein (HsKin17 protein) (KIN, antigenic determinant of recA protein homolog) (Mouse).
GN Name=KIN; Synonyms=Kin17; ORFNames=RP11-264C14.1-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Testis;
RX MEDLINE=20419742; PubMed=10964102; DOI=10.1093/carcin/21.9.1701;
RA Kannoche P., Maufrey P., Pinon-Lataillade G., Mattei M.G.,
RA Sarsain A., Daya-Grosjean L., Angulo J.F.;
RT "Molecular cloning and characterization of the human KIN17 cDNA encoding a component of the UVC response that is conserved among metazoans.";
RT Carcinogenesis 21:1701-1710(2000).
```


Db 121 LGNGCHVADTDKGFYIYDRDPETLAMOQKWKQKMDKDAERLAEFFIEQVRRGK 180
 Qy 182 EGKEQTPVFTLSRNEEBKVTNKNKGAGGAGATTSSKSLGPSAL----KLGSAAS 238
 Db 181 TEEBPTSGVSELKRENEEDTIKELKLG-----SKQOQSTPSAVISKEPFDALDD 231
 Qy 239 GKREKSSQSQAPAKKK-SALDEIMELEBEKK-RTARTDAWLQPGVVKVVKIITKLGEKY 296
 Db 232 GKKEKKIKATSNGETKKSALDELQEEQKKEKNRKYWLAEGLVVKLISRSLOEKY 291
 Qy 297 HKKGGVVKVEIDRYTAVVKNMDSGRDLKLDQTHLETVIPAGKRVLVNLNGYGRNEGTL 356
 Db 292 YKEGVVVEIEKYRAKIKLLETGERKLKVDQALHETVIPAGVQILVNLNGYRGCTAVLK 351
 Qy 357 SINEKAFSATIVTETGLKRRVEGQYEDISKL 390
 Db 352 AINTERYSVTETASGLKGLVSNVAYEDISKL 385

RESULT 7

Q9CV58 MOUSE
 ID Q9CV58 MOUSE PRELIMINARY; PRT; 195 AA.
 AC Q9CV58;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DE Mus musculus adult male tongue cDNA, RIKEN full-length enriched
 DE library, clone:2310020017 product:antigenic determinant of rec-A
 DE protein, full insert sequence. (Fragment).
 GN Name=Kin;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning."
 RL Meth. Enzymol. 303:19-44(1999).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barch G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Guetincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 RN [3]

NUCLEOTIDE SEQUENCE.

STRAIN=C57BL/6J; TISSUE=Tongue;

the FANTOM Consortium;

the RIKEN Genome Exploration Research Group Phase I & II Team;

"Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes."
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer."
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tanaka M., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK009429; BAB26281.2; -, mRNA.
 DR MGI; MGI:96676; Kin.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR007087; Znf_C2H2.
 DR SMART; SM00355; Znf_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 FT NON TER 195 195
 SQ SEQUENCE 195 AA; 23518 MW; 7E112E9CA820BFA6 CRC64;

Query Match

Best Local Similarity 52.1%; Score 1056; DB 2; Length 195;

Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MGKSDFLSPKAIANRIKSLGKLRLWYQCMQKQCRDENGFKCHCMSESHORQLLAS 60
 Db 1 MGKSDFLSPKAIANRIKSLGKLRLWYQCMQKQCRDENGFKCHCMSESHORQLLAS 60
 Qy 61 PQQPMDFSEBEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDP 120
 Db 61 PQQPMDFSEBEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDP 120
 Qy 121 WLGRGELCKVDETPKGYIQVDRDPETIRRLQLEKKKKQDLDEEKTAKFIEQVRG 180
 Db 121 WLGRGELCKVDETPKGYIQVDRDPETIRRLQLEKKKKQDLDEEKTAKFIEQVRG 180
 Qy 181 LEGKEQETPVFTELS 195
 Db 181 LEGKEQETPVFTELS 195

RESULT 8

Q9VPH4 DROME
ID Q9VPH4 DROME PRELIMINARY; PRT; 390 AA.
AC Q9VPH4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2004 (TrEMBLrel. 26, Last annotation update)
DE CG5649-PA.
GN Name=kin17; ORFName=CG5649;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=107311132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobbart J.C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington I., Simpson M., Skupski M.P., Smith T.,
RA Shue B.C., Siden-Kianos I., Saunders R.D.C., Scheeler F., Shen H.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426065; PubMed=125373568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svitek R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svitek R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [5]
RN NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svitek R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RN NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
CC -!- INTERACTION:
CC Q9V116:CG1041; NbExp=1; IntAct=EBI-193999, EBI-175006;
CC Q9VPS5:Hsp60B; NbExp=1; IntAct=EBI-193999, EBI-159512;
CC P21187:pAbp; NbExp=1; IntAct=EBI-193999, EBI-103658;
CC EMBL; AE003591; AAF51578.1; -, Genomic DNA.
DR IntAct; Q9VPH4; -;
DR Ensembl; CG5649; Drosophila melanogaster.
DR FlyBase; FBgn0024887; CG5649.
DR FlyBase; FBgn0024887; kin17.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; ZnF_C2H2.
DR SMART; SM00355; ZnF_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ SEQUENCE 390 AA; 45355 MW; 5F8231706C2D431F CRC64;
Query Match 51.9%; Score 1052.5; DB 2; Length 390;
Best Local Similarity 52.2%; Pred. No. 9.4e-55;
Matches 206; Conservative 69; Mismatches 107; Indels 13; Gaps 5;
QY 1 MKSDFLSPKAIANRIKSKGLQKRWYCOMCQKQCRDENGFKCHCMSESHORQLLASN 60
DB 1 MGRAEVGTPKYLANKWKSKGLQKRWYCOMCEKQCRDENGFKCHTWSHQLLLFADN 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVVNYEISHREHIMNATOWETLTDFTK 120
DB 61 FQKFLHSFKSESDGCMELLRRRFGTKTSANKIYQEVIAHKEHIMNATRLWLTSLDYVK 120
QY 121 WLGREGLCVDETPKGWYQIVDRDPETIRRQLEKLEKKKKQDLDEEKTAKTIEQVR-- 178
DB 121 WLGRTGQVIADETEGWFWYVDRSPFAMERQAKADRKEMKEDDEERMADEIEQIKNA 180
QY 179 RLEGEKEQE-TPVFTFLSR-ENEEKVTFLNKNKGAGGAGATTSSKSSSIPSPALKLGS 236
DB 181 KAKDGEEDGQEKFTLKEENEPLKDIRLEK-----KFQPDVTVLGKSALAKRAP 232
QY 237 ASGKKESSQSSAQPAKKKKSALDEIMELEBKK-RTARTDAWLQPGIVVKITKLGK 295
DB 233 EAEKVFKKPKSVAGDSQTRSVLDIIKKQESKKEANRKYWLHKGIIVVKISKMGK 292
QY 296 YHKKGKGVKEIDRVYAVVQMTDSGRKLKLDOTHLETVIPAGPKRVLVNGGYRNEGTL 355
DB 293 FPKQKAVLDVDRYQGGIKFLETGKLVQDAHLETVIPALDKPVVNVNGYARGSEALL 352


```
QY 181 LE-----GKEQETPVPTELSRENEE-----KYTFNLNKGAGGSAGATTSSK 222
Db 181 ASKLANGGEGEGTSGNDEVDGDDERKDKEDLRLKSGVKVGFGAL---GGYKQVATGKB 237
QY 223 SSLGFSALKLILGSAAS-----GKRKSSQSAQPAKKKKSALDEIMLEEEKK-RTARTD 276
Db 238 R--GESSKLLFDEENDKVEREKRKRSKSGDSRSEKERSALDELMKEEEKKERNRKO 295
QY 277 AWLQPGIVVKIITKLGKEK-YHKHKGWVKEVDRTYAVVKTDSGDRCLKDQTHLETVIP 335
Db 296 YWLFEGIIKVMKSKALAEKGYKQKGVKVKVDNYVGEIKMLDSKHLVRVDQKELETVP 355
QY 336 ARKRVVLNCGYRGNEGTLSEINEKAFSATIVETGTPKLGRRVEGIQYEDISKLA 391
Db 356 QIGGMVKVINGAYRGSNARLLGVDTEKFCARQVIEKGVVDGRVIRKSIYEDICKLA 411

RESULT 11
Q9XWF2_CAEEL PRELIMINARY; PRT; 404 AA.
AC Q9XWF2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein Y52B11A.9.
GN ORFNames=Y52B11A.9;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1];
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=9069613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; AL032654; CAA21720.1; -; Genomic_DNA.
DR PIR; T27106; T27106.
DR WormBase; WBGene00013128; Y52B11A.9.
DR Ensembl; Y52B11A.9; Caenorhabditis elegans.
DR WormPep; Y52B11A.9; CE20297.
DR GO; GO:0003634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:007275; P:development; IEA.
DR InterPro; IPR002358; Ribosomal_L6_1.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00525; RIBOSOMAL_L6_1; UNKNOWN 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 404 AA; 46401 MW; 9F55A1413525F057 CRC64;

Query Match 47.8%; Score 969.5; DB 2; Length 404;
Best Local Similarity 48.3%; Pred. No. 8.7e-50;
Matches 199; Conservative 75; Mismatches 109; Indels 29; Gaps 8;

QY 1 MGKSDFLSPKATNRKSGLOKLRVYCOMCQKQCDENGFKCHCWSESHORQLLASFN 60
Db 1 MGKHEKSGSKDLNARKTSGLOKLRVYCOMCQKQCDANGFKCHLTSSEAHORQLLFAFN 60

QY 61 PQQFMDFSEEFNDLLELRRRFGKRVHNNIYVNEYISHREIHIMNATOWETLTDFTK 120
Db 61 SNSYLRFQFNDPEKFNFWLLRTSYGKRVANEVYNFADKGVHVNSTVWHSLTGFVQ 120

QY 121 WLGREGLCKVDETPKGWYQYIDRDPETTRRQLELEKKKKQDLDDSEKTAFTIEVRG 180
Db 121 YLGSSGCKKIDEGKGYAYIDQ--EALIRKEEDQRQKQKDDDERHMQMDGMVQRG 178

QY 181 --LEGKEQETPVPTELSRENEEKTFTNLNKG-----AGGSAGATTSSKSSL----- 225
```

```
Db 179 KELAGDDHEYEATELIRDTPDQKIQLDNLGILDRKLDVKSGVASAKISIFDMPKVKE 238
QY 236 -----GPSALKLLGSAAGKRESQSAQPAKKKKSALDIMELEEBKK-RTARTDAWL 279
Db 239 DPDEPGPSQ-----PSRSGKRSRSPAAKFKSKSALDEIKEMEERKKERKNRKYWM 294
QY 280 OPGIWKIITKLGKEKHKKGVVKEVDRTYAVVKTDSGDRCLKDQTHLETVIPAPGK 339
Db 295 REGIVKVVITKSLGSEYYKAGWVRKVVDDYTAQVKL--DDGTVVKLDQEHVETVPSLR 353
QY 340 RVLNCGYRGNEGTLSEINEKAFSATIVETGTPKLGRRVEGIQYEDISKLA 391
Db 354 QMMIVNGAYRQEQATLESIDEKFRSLRLKIASGPTGRGQID--VPYEDASKLA 404

RESULT 12
Q4TDV9_TETNG PRELIMINARY; PRT; 269 AA.
AC Q4TDV9_TETNG;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF6005, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG0002601001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
Anthonard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
Wincker P., Lander E.S., Weissbach J., Roest Crollius H.;
RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
the early vertebrate proto-karyotype.";
RL Nature 431:946-957 (2004).
RN [2];
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
CC EMBL; CAEE01006005; CAF88923.1; -; Genomic_DNA.
DR NON_TER 1
FT NON_TER 269
SQ SEQUENCE 269 AA; 31736 MW; E6883839A2CD17AF CRC64;

Query Match 46.0%; Score 932.5; DB 2; Length 269;
Best Local Similarity 68.4%; Pred. No. 8.6e-48;
Matches 186; Conservative 33; Mismatches 38; Indels 15; Gaps 5;

QY 55 LLAENPQQFMDYFSEEFNDLLELRRRFGKRVHNNIYVNEYISHREIHIMNATOWET 114
Db 1 LLAENPNRFNDHFQEFKQFLELRRRFGKRVHNNIYVNEYISDRQIHIMNATRWET 60

QY 115 LTDFTKWLGREGLCKVDETPKGWYQYIDRDPETTRRQLELEKKKKQDLDDSEKTAFTIE 174
Db 61 LTDFTKWLGREGLCKVDETPKGWYQYIDRDPETTRRQLELARKKKHDLDDSEKTAFTIE 120

QY 175 EQVRGLEGKE-QETPVPTELSRENEEKTFTNLNKGAGGSAGATTSSKSLGPSALKLL 233
```

Db 121 EQVRGRDRETETPLTYELKRENEEKKVAFNL--GASLCVAGPSKSPAVGVGALKKE 178
QY 234 GSAAGKRKESSOQAQPAKKKKSALDEIMELEBEEKKR-----TARTDAWLOPGIIVKLIYK 290
Db 179 AAAAASVKRKDAS--SESKTEKKRKSALSALEIEMEKEKKRQQOPLRTDYWLQPNIVVVKIYK 237
QY 291 KLGEKVHKKGVVKEVIDRYTAVVKTDSGDR 322
Db 238 KLGERVHKRGV-----MVRPRDSDVDR 261

RESULT 13

Q75LU5_ORYSA
ID Q75LU5_ORYSA PRELIMINARY; PRT; 430 AA.
AC Q75LU5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein OSJNB0015121.3.
GN Name=OSJNB0015121.3;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Buell C.R., Yuan Q., Qiyang S., Liu J., Gansberger K., Jones K.M.,
RA Overton I.F., Taitrin T., Kim M.M., Bera J.J., Jin S.S.,
RA Padrosh D.W., Tallon L.J., Koo H., Zismann V., Hsiao J.J., Blunt S.,
RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,
RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
RA White O., Salzberg S.L., Fraser C.M.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Buell R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC091302; AAR0634.1; -; Genomic_DNA.
DR Gramene; Q75LU5; -;
DR InterPro; IPR005824; KOW.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00467; KOW; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 430 AA; 49067 MW; 6151D748ED2F4690 CRC64;

Query Match 45.0%; Score 913; DB 2; Length 430;
Best Local Similarity 44.7%; Pred. No. 2.2e-46;
Matches 192; Conservative 73; Mismatches 123; Indels 42; Gaps 7;
QY 1 MGKSDFLSPKAIANRIKSGLOKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFOQA 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVVNEVISHREHIMNATOWETLTDTFK 120
Db 61 PDRVEGSEEFDAFLTLRRRAHRSRIAATVTVNNEFIADRHVHNMNSTRWATLTFVK 120
QY 121 WLGRGLCKVDTPKGVWIOYIDRDPETIRROLEKKKQDLDDEEKTAFIEQVRR- 179
Db 121 FLGRGHCKVEDTPKGVWFIYIDRSEQAVKARLKRKIKSDLAEDERQERMIARQIERA 180
QY 180 -----GLEGEQETVPVTELSRENE-EKVTFLNKGAGGSAGATTS 220
Db 181 QOSMGKTNGELGDDASPDGSESGSESEYSDSENHGEQEDAEKANKAAGKIAIAL-- 238
QY 221 KSSSLGPSALKLIGS-----AASGRKE-----SSOSSAPAKKKKSGALDEI 262
Db 239 QRAVFGPKVNLDDPKPKVFGFEEDEVSARDKEELAKKGGKDAINAAEARRSALDEL 298
QY 263 MELEBE-KKRTARTDAWLOPGIIVKLIYKLGK-YHKKGVVKEVIDRYTAVVKTDSG 320

Db 299 MKEEKAKERSNRKDYWLCPGIWVKMSKSLAEKGYCKQKGVVVKRVIDKYVGEIEMLESK 358
QY 321 DRILKLDQTHLETVIPAGKRVLVNGYGRNGTLESINEKAFSAFATIVTETGPLKGRRYE 380
Db 359 HVLRVDDQDELETVIPQIGGLVIVNGAYGSGNARLLSVTERFCARVQVEKGLYDGKVLK 418
QY 381 GIQYEDISKL 390
Db 419 AIEYEDICKI 428

RESULT 14

Q9AY87_ORYSA
ID Q9AY87_ORYSA PRELIMINARY; PRT; 402 AA.
AC Q9AY87;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein OSJNB0004B24.17.
GN Name=OSJNB0004B24.17;
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,
RA Zismann V., Pal G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Buell R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC084319; AAG59654.1; -; Genomic_DNA.
DR Gramene; Q9AY87; -;
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; ZNF_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 402 AA; 45962 MW; A1F4E38ECC60D13E CRC64;

Query Match 41.4%; Score 839; DB 2; Length 402;
Best Local Similarity 42.3%; Pred. No. 5.2e-42;
Matches 182; Conservative 67; Mismatches 111; Indels 70; Gaps 8;
QY 1 MGKSDFLSPKAIANRIKSGLOKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFOQA 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVVNEVISHREHIMNATOWETLTDTFK 120
Db 61 PDRVEGSEEFDAFLTLRRRAHRSRIAATVTVNNEFIADRHVHNMNSTRWATLTFVK 120
QY 121 WLGRGLCKVDTPKGVWIOYIDRDPETIRROLEKKKQDLDDEEKTAFIEQVRR- 179
Db 121 FLGRGHCKVEDTPKGVWFIYIDRSEQAVKARLKRKIKSDLAEDERQERMIARQIERA 180
QY 180 -----GLEGEQETVPVTELSRENE-EKVTFLNKGAGGSAGATTS 220
Db 181 QOSMGKTNGELGDDASPDGSESGSESEYSDSENHGEQEDAEKANKAAGKIAIAL-- 238
QY 221 KSSSLGPSALKLIGS-----AASGRKE-----SSOSSAPAKKKKSGALDEI 262
Db 239 QRAVFGPKVNLDDPKPKVFGFEEDEVSARDKEELAKKGGKDAINAAEARRSALDEL 298

Qy	263	MELBEE-KKRTARTDAMIQPGIVVKIIITKKLGEK-YHKKKGVVKEVIDRVTAVVKMTDSG	320
Db	299	MKEEEKAKERSNRKQYWLCPGLVVKVMSKSLAEKGYCKQKGVKRVVDKXVGEIEMLESK	358
Qy	321	DRILKLDQTHLETVIPAPGRKRVLVNLGGYRGNGEGTLESINEKAFSATIVETGTPLKGRRVE	380
Db	359	HVLRVDDQDELTIVIQGG-----LVEKGLYDGKVLK	390
Qy	381	GIQYEDISKL	390
Db	391	AIYEYEDICKI	400

RESULT 15

Q60L39_CAEBR

ID	Q60L39_CAEBR	PRELIMINARY;	PRT;	397	AA.
AC	Q60L39;				
DT	25-OCT-2004	(TREMBLrel. 28, Created)			
DT	25-OCT-2004	(TREMBLrel. 28, Last sequence update)			
DT	25-OCT-2004	(TREMBLrel. 28, Last annotation update)			
DE	Hypothetical protein CBG23783.				
GN	Names=CBG23783;				
OS	Caenorhabditis briggsae.				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;				
OX	Rhabditidae; Peloderinae; Caenorhabditis.				
NCBI_TaxID=6238;					
[1]	RN				
RP	NUCLEOTIDE SEQUENCE.				
RG	The C.briggsae Sequencing Consortium;				
RL	Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.				
CC	-!- CAUTION: The sequence shown here is derived from an				
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is				
CC	preliminary data.				
DR	EMBL; CAAC01000183; CAE56163.1; -; Genomic_DNA.				
DR	InterPro; IPR007087; Znf_C2H2.				
DR	ProSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.				
DR	Hypothetical protein.				
SK	SEQUENCE 397 AA; 45371 MW; E5671B0DFE2F9237 CRC64;				

Query Match	41.2%;	Score 835;	DB 2;	Length 397;
Best Local Similarity	44.7%;	Pred. No. 8.9e-42;		
Matches 183;	Conservative 65;	Mismatches 131;	Indels 30;	Gaps 10;
QY	1	MGKSDFLSPKAIANRIKSGLOKLRYWCQKOCORCDENGFKCHCMSESHORQLLASEN	60	
DB	1	MGKHEKSSKDLANRTKSGLOKLKWFQCMQKQCRDANGFKCHLTSEAKORQLLPAEN	60	
QY	61	PQOFMDYFSEEFNRNDFLELLRRRFGTKRVHNNIVNYEYISHREHHNATQWETLTDFTK	120	
DB	61	SNSYLR--XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGXGHVMNSTVHSLTGFVQ	118	
QY	121	WLREGGLCKVDETPKGWYIOYDRDPETIRQLELEKKKODLDEKTAKFTEEQVRG	180	
DB	119	YLSSGCKKIDEGKGYIAYIDQ--EAVIRKEEBQRKQOERDEDRHQLIVEEMVARG	176	
QY	181	LEKGEQETPVF--TELSRENEEEKVTENL-----NKGAGGSAGATT-----SKSSS	224	
DB	177	RELAGEDTVDEATELRREDPDQKIQLDLVGFQKPKQNDQFASKTNSIFDLKPVKXEDP	236	
QY	225	LGSALKLGSASAGKRKESQSSAQPAKK-KGSALDEIMELEBEKK-RTARTDAWLQPG	282	
DB	237	DGP-----GPSGAPKRRRSRSPKSSKNAKKSALDEIKEMESRKKRNKRDYNNREG	290	
QY	283	IVVKIITKLGEKYHKKGVVKEVIDRYTAVKMTDSDGRLKLDOTHELTVIPAPGRVL	342	
DB	291	IVVKVTYKLSGSQYYSKGVVKWIDYTAQVKL--DDGTVVKLQDEHVEVTVISGRTML	349	
QY	343	VLNGVYRGNEGTLLESINEKAFSATIVIETGPLKGRRVEGIQYEDISKLA	391	
DB	350	VNMGAYRGQATLESIDEXHFSURLKIASGPTRGROID-VPYEDASQWA	397	

Search completed: November 23, 2005, 16:33:32

Job time : 155.61 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:23:48 ; Search time 33.9133 Seconds
(without alignments)
953.202 Million cell updates/sec

Title: US-09-555-529-25

Perfect score: 2028

Sequence: 1 MGKSDPLSPKAIANRIKSG.....GPIKGRVGEIQVEDISKLA 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5 COMB.pap.*
- 2: /cgn2_6/ptodata/1/iaa/6 COMB.pap.*
- 3: /cgn2_6/ptodata/1/iaa/H COMB.pap.*
- 4: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pap.*
- 5: /cgn2_6/ptodata/1/iaa/RE COMB.pap.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	375.5	18.5	136	2	US-09-270-767-32008
2	348.5	17.2	274	2	US-09-248-796A-20041
3	124	6.1	2349	2	US-09-538-092-914
4	122	6.0	1564	2	US-10-144-198-2
5	122	6.0	1564	2	US-10-144-198-4
6	121.5	6.0	2476	2	US-09-824-574-7
7	121	6.0	558	1	US-08-285-440-6
8	121	6.0	558	1	US-08-630-349-6
9	119.5	5.9	1676	2	US-09-949-016-7610
10	119	5.9	810	2	US-09-248-796A-20281
11	118.5	5.8	1507	2	US-08-929-329-5
12	117	5.8	484	2	US-09-949-016-8999
13	117	5.8	532	1	US-08-285-440-5
14	117	5.8	532	1	US-08-630-349-5
15	117	5.8	538	2	US-09-949-016-6479
16	117	5.8	541	2	US-09-949-016-7217
17	116.5	5.7	683	6	5210183-3
18	116.5	5.7	845	2	US-10-104-047-2309
19	116	5.7	344	6	5210183-2
20	114	5.6	440	2	US-09-762-194-2
21	114	5.6	440	2	US-09-762-194-4
22	113	5.6	1298	2	US-09-645-456A-14
23	113	5.6	1298	2	US-09-425-324A-14
24	113	5.6	1298	2	US-09-645-791-14
25	113	5.6	1394	2	US-09-902-540-16497
26	113	5.6	1960	2	US-09-538-092-1077
27	112	5.5	793	2	US-09-538-092-1271

```

28      112      5.5      1960      2      US-09-949-016-10872      Sequence 10872, A
29      111      5.4      1427      2      US-09-538-092-1044      Sequence 1044, Ap
30      110.5      5.4      847      2      US-09-976-594-401      Sequence 401, App
31      110.5      5.4      853      2      US-09-949-016-11423      Sequence 11423, A
32      110.5      5.4      2186      2      US-09-949-016-10828      Sequence 10828, A
33      110      5.4      723      1      US-07-814-964-11        Sequence 11, Appl
34      110      5.4      723      1      US-08-358-442-11        Sequence 11, Appl
35      110      5.4      723      1      US-08-328-809-6         Sequence 6, Appl
36      110      5.4      723      1      US-08-866-840-6         Sequence 6, Appl
37      110      5.4      723      4      PCT-US92-11107-11        Sequence 11, Appl
38      110      5.4      1055      2      US-09-949-016-9776      Sequence 9776, Ap
39      110      5.4      1306      2      US-09-645-456A-10        Sequence 10, Appl
40      110      5.4      1306      2      US-09-425-324A-10        Sequence 10, Appl
41      110      5.4      1306      2      US-09-645-791-10        Sequence 35, Appl
42      109      5.4      440      2      US-08-302-756B-35        Sequence 13787, A
43      109      5.4      515      2      US-09-248-796A-19787      Sequence 2753, Ap
44      109      5.4      718      2      US-09-540-236-2753      Sequence 6779, Ap
45      109      5.4      994      2      US-09-949-016-6779

```

ALIGNMENTS

RESULT 1

US-09-270-767-32008
; Sequence 32008, Application US/09270767
; Patent No. 6703491

; GENERAL INFORMATION:

; APPLICANT: Homburger et al.

; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*

; FILE REFERENCE: File Reference: 7326-094

; CURRENT APPLICATION NUMBER: US/09/270,767

; CURRENT FILING DATE: 1999-03-17

; NUMBER OF SEQ ID NOS: 62517

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 32008

; LENGTH: 136

; TYPE: PRT

; ORGANISM: *Drosophila melanogaster*

US-09-270-767-32008

Query Match 18.5%; Score 375.5; DB 2; Length 136;

Best Local Similarity 55.6%; Pred. No. 7.6e-27;

Matches 74; Conservative 24; Mismatches 34; Indels 1; Gaps 1;

```

Qy      259      LDIMELESEKK-RTARTDAWLPQIVVKKIITKKLGEYHKYKGVKVEIDRYTAVVKMT 317
Db      1      LDEIIKQESKSKERANKRKYMLHKGIYVVKFISKSMGSKFFKQKAVLDVIDRYOGKIKFL 60

Qy      318      DSGDRKLKLDTHLETVIPAPGKRVLVNLGGYRGNEGTLSEINEKAFSATIVETGPKLGR 377
Db      61      ETEGKLKVDQALETVIPALDKPVMVNGAYRGSEALLKLDERRYSVSVELHGLKGR 120

Qy      378      RVEGIQYVEDISKL 390
Db      121      IVDNVQYVEDISKL 133

```

RESULT 2

US-09-248-796A-20041
; Sequence 20041, Application US/09248796A
; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICA

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13


```

; ORGANISM: Homo sapiens
US-10-144-198-4
Query Match      6.0%; Score 122; DB 2; Length 1564;
Best Local Similarity 24.2%; Pred. No. 0.077;
Matches 66; Conservative 38; Mismatches 91; Indels 78; Gaps 13;

QY 116 TDTFTKWLREGGLCKVDET-----PKG-WIYQIDRDPEITRRQL-ELEKKKKQDLDD 165
DB 40 TQCRNWL-KTGNCLYNTCFVHGSPRGKGYSSNRYRSPERTGDLRMRMKNKRDQDVT 98

QY 166 E-----EKTAFIEQVRGLEKEQETPVFTLSRENEEEKVTFLNKNKGAGGSAG--- 216
DB 99 EPQKRNTESSSPVKRESSGRHREKEDIKITERPESSEENVENVETNRDSDNGDINY 158

QY 217 -----ATTSSSLGPSALKLLGSA 237
DB 159 DYVHELSELMKROKIQRELMKLEQENNEKREIIKKVEPSVVRVSKLSPSP-SLRKSSK 217

QY 238 SGRKESOSSAQPAKKKSA-----LDE--IMELEEEKKRTARTDAMLQPGIVVKIIT 289
DB 218 SPKRKSPKSSSSAKDKRTSAVSPLLDQQRNSKTNQSKKGPRTPS---PPPIPEDIA 275

QY 290 KKLGEKYHKKGWVVDYRTAVVWMTDSGDR 322
DB 276 --LGKXY-KEKYVKDRIEB-----KTRDGKDR 300

RESULT 6
US-09-824-574-7
; Sequence 7, Application US/09824574
; Patent No. 6693170
; GENERAL INFORMATION:
; APPLICANT: Rouleau, Natalie
; APPLICANT: Moilanen, Anu-Maarit
; APPLICANT: Palvimo, Jorma J.
; APPLICANT: Jone, Olli A.
; TITLE OF INVENTION: ARIp4 Gene and Protein
; FILE REFERENCE: 2630-109
; CURRENT APPLICATION NUMBER: US/09/824,574
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 7
; LENGTH: 2476
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-824-574-7

Query Match      6.0%; Score 121.5; DB 2; Length 2476;
Best Local Similarity 22.8%; Pred. No. 0.17;
Matches 68; Conservative 41; Mismatches 108; Indels 81; Gaps 13;

QY 122 LGRGLCKVDETPKGWYQIYDRDP-----ETIRROLE--LEKKKKQDLDDDEKTKAK 171
DB 247 LGRKELSTIMDENQWYCYIQPEPLDLVTACNSVENLEQLQONKKIKVDSEKTSK 306

QY 172 FIEQVRVRLGKEQETPVFTLSRENEEEKVTFLNKNKGAGGSAGATTSSK----- 222
DB 307 VCDQ-----TSKFPSPKSSSCNGEKK-----LESCSGSVSYSHSALSVPKEMI 354

QY 223 -----SSLGPSALKLGSAAGKREKSSOSSAQPAKKKKSALDEI-----MELEEE 268
DB 355 KKTTLKETTSNNSSVIKFLQADNSSEMTSAMKLCQ-LKSPKSVLDDIKKAHLEED 413

QY 269 KRTARTDAMLQPGIVVKIITKLGEKYHKKGWVVDYRTAVVWMTDSGDRKLDDOT 328
DB 414 LNSEIQ-----LDDVHKEKN-----TKDLKSTDAKSTKLKGK 447

QY 329 HLETVIPAPGRVILNGVGRNEGTLSEINEKAFSAFIVTETGLKGR-RVEGIQYE 385
DB 448 --EKSYSYTERKEEFLKLOA--RSSVKAIDGEEQRAHKST----SCEHKGSGRKGDSQYE 497
```

```

RESULT 7
US-08-285-440-6
; Sequence 6, Application US/08285440
; Patent No. 5532337
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/285,440
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 558 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; LIBRARY:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
```


QY 266 EBEKRTARTDAWLQBGIVVWIKITKLGKGYHKHKKGVVKE 305
Db 1254 EBEKRT-----IEAEKKABEERKRIEAVKK 1278

RESULT 12
US-09-949-016-8999
; Sequence 8999, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8999
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8999

Query Match 5.8%; Score 117; DB 2; Length 484;
Best Local Similarity 20.2%; Pred. No. 0.039;
Matches 67; Conservative 48; Mismatches 93; Indels 124; Gaps 14;

QY 48 ESHQROLLASENPOQFMDYFSEEFNDFLELLRRRFGTKRVHNNIVVNEYISHREHIHM 107
Db 108 ERRQKRLQALERQKEFDFTTDASUS-----LPSSRMQNDTAENETTEKEEKSES 158
QY 108 NATQWETLTDFTKWLREGGLCKVDETPKGWYIQYIDRDPETIRRQJ-----ELEKKKQ 161
Db 159 RQERYE-----IEET-----ETVTKSYQKNDWRDAENKKE 189
QY 162 DLDEEKTAKFEEQVRRLEG-----KEQETPV-----FTLSRENEE 200
Db 190 DKEKEEEE---EEKPKRGSIGENQIKDEKIKKDEKPEEVSFMDRKKGFTVKSQNGE 245
QY 201 -----EKVTFNLKGGAGATTSSSLGPSALKLLGSAASGKR----- 241
Db 246 FMTHKLKHTENTFSRPGGRASVDTKAEG-----APQVEAGKRLBELRRRRRGTESE 297
QY 242 -----KESQSSA---QPAKKKSALDEIMLEEKKRTARTDAWLQPGIVVWIKITKLG 293
Db 298 EPEKLKQKQEALELEULKKEERKRVLEEEQQRKQEAD-----RKL 344
QY 294 EKYHKKGKGVVKEVIDRYTAVV-----KMTDSG 320
Db 345 EBEKRR--LKEBIERRRAEAEKQKMPEDG 374

RESULT 13
US-08-285-440-5
; Sequence 5, Application US/08285440
; Patent No. 5532337
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.

Query Match 5.8%; Score 117; DB 1; Length 532;
Best Local Similarity 20.2%; Pred. No. 0.045;
Matches 67; Conservative 48; Mismatches 93; Indels 124; Gaps 14;

; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: DisplayWrite
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/285,440
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM:
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; IMMEDIATE SOURCE:
; CLONE:
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT:
; MAP POSITION:
; UNITS:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; AUTHORS:
; TITLE:
; JOURNAL:
; VOLUME:
; ISSUE:
; PAGES:
; DATE:
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-08-285-440-5

Query Match 5.8%; Score 117; DB 1; Length 532;
Best Local Similarity 20.2%; Pred. No. 0.045;
Matches 67; Conservative 48; Mismatches 93; Indels 124; Gaps 14;
QY 48 ESHQROLLASENPOQFMDYFSEEFNDFLELLRRRFGTKRVHNNIVVNEYISHREHIHM 107

Db 99 ERQKRLQEALEKQKFDPTTIDASLS-----LPSRRMQNDTAENETTEKEEKSES 149
Qy 108 NATOWETLDTFTKWLREGGLCKVDETPKGWYIOYIDRDPETIRROL-----ELEKKKKQ 161
Db 150 RQERYE-----ETVTKSYQKNDWRDAEENKKE 180
Qy 162 DLDEEKTAKFIEQVRRGLEG-----KEQETPV-----FTLSRENEE 200
Db 181 DKEKEEEE-----BEKPKRGSGIGENQIKDEKIKKDKPEKVEVKSFMDRKKGFTVKSQNGE 236
Qy 201 -----EKVTFNLNKGAGSGAGATTSSKSSSLGPSALKLLGSAASGKR----- 241
Db 237 FMTHKLKHTENTSRPGGRASVDTKEAG-----APQVEAGKRLLEELRRRGETESE 288
Qy 242 -----KESSOSSA---OPAKKKKSALDEIMELEBEKKRTARTDAWLOPGIVVVIITKKLG 293
Db 289 EPEKLKQKQQAALAELEELKKREERRKVLSEEEQRKQBEAD-----RKLR 335
Qy 294 EKYHKKKGVVKEVIDRYTAVV-----KMTDSG 320
Db 336 EBEKKR--LKEEIERRRRAEAEKROKMPEDG 365

RESULT 14

US-08-630-349-5
; Sequence 5, Application US/08630349
; Patent No. 5739008
; GENERAL INFORMATION:
; APPLICANT: Ken'ichiro HAYASHI et al.
; TITLE OF INVENTION: POLYPEPTIDE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 144 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,349
; FILING DATE: April 10, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/285,440
; FILING DATE: August 4, 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/858,947
; FILING DATE: March 27, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-630-349-5

Query Match 5.8%; Score 117; DB 1; Length 532;
Best Local Similarity 20.2%; Pred. No. 0.045;
Matches 67; Conservative 48; Mismatches 93; Indels 124; Gaps 14;
Qy 48 ESHQRQLLASNPQOFMDYFSEFRNDFLELLRRRFGTKRVHNNIVVYVISHREIHM 107
Db 99 ERQKRLQEALEKQKFDPTTIDASLS-----LPSRRMQNDTAENETTEKEEKSES 149

Qy 108 NATOWETLDTFTKWLREGGLCKVDETPKGWYIOYIDRDPETIRROL-----ELEKKKKQ 161
Db 150 RQERYE-----ETVTKSYQKNDWRDAEENKKE 180
Qy 162 DLDEEKTAKFIEQVRRGLEG-----KEQETPV-----FTLSRENEE 200
Db 181 DKEKEEEE-----BEKPKRGSGIGENQIKDEKIKKDKPEKVEVKSFMDRKKGFTVKSQNGE 236
Qy 201 -----EKVTFNLNKGAGSGAGATTSSKSSSLGPSALKLLGSAASGKR----- 241
Db 237 FMTHKLKHTENTSRPGGRASVDTKEAG-----APQVEAGKRLLEELRRRGETESE 288
Qy 242 -----KESSOSSA---OPAKKKKSALDEIMELEBEKKRTARTDAWLOPGIVVVIITKKLG 293
Db 289 EPEKLKQKQQAALAELEELKKREERRKVLSEEEQRKQBEAD-----RKLR 335
Qy 294 EKYHKKKGVVKEVIDRYTAVV-----KMTDSG 320
Db 336 EBEKKR--LKEEIERRRRAEAEKROKMPEDG 365

RESULT 15

US-09-949-016-6479
; Sequence 6479, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6479
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6479

Query Match 5.8%; Score 117; DB 2; Length 538;
Best Local Similarity 20.2%; Pred. No. 0.045;
Matches 67; Conservative 48; Mismatches 93; Indels 124; Gaps 14;
Qy 48 ESHQRQLLASNPQOFMDYFSEFRNDFLELLRRRFGTKRVHNNIVVYVISHREIHM 107
Db 105 ERQKRLQEALEKQKFDPTTIDASLS-----LPSRRMQNDTAENETTEKEEKSES 155
Qy 108 NATOWETLDTFTKWLREGGLCKVDETPKGWYIOYIDRDPETIRROL-----ELEKKKKQ 161
Db 156 RQERYE-----ETVTKSYQKNDWRDAEENKKE 186
Qy 162 DLDEEKTAKFIEQVRRGLEG-----KEQETPV-----FTLSRENEE 200
Db 187 DKEKEEEE-----BEKPKRGSGIGENQIKDEKIKKDKPEKVEVKSFMDRKKGFTVKSQNGE 242
Qy 201 -----EKVTFNLNKGAGSGAGATTSSKSSSLGPSALKLLGSAASGKR----- 241
Db 243 FMTHKLKHTENTSRPGGRASVDTKEAG-----APQVEAGKRLLEELRRRGETESE 294
Qy 242 -----KESSOSSA---OPAKKKKSALDEIMELEBEKKRTARTDAWLOPGIVVVIITKKLG 293
Db 295 EPEKLKQKQQAALAELEELKKREERRKVLSEEEQRKQBEAD-----RKLR 341
Qy 294 EKYHKKKGVVKEVIDRYTAVV-----KMTDSG 320

Db 342 EEEKER--LKBEIERRRAAEKROKMPEDG 371

Search completed: November 23, 2005, 16:35:41
Job time : 35.9133 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:24:04 ; Search time 113.709 Seconds
(without alignments)
1436.746 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 2028
Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPKGRVBEIGQVEDISKLA 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1895	93.4	393	4	US-10-408-765A-1465
2	1052.5	51.9	390	6	US-11-097-143-41040
3	913	45.0	437	4	US-10-425-114-48183
4	913	45.0	437	4	US-10-425-114-58982
5	908	44.8	424	4	US-10-425-115-341788
6	901.5	44.5	423	5	US-10-739-930-8499
7	613	30.2	176	4	US-10-424-599-168326
8	531	26.2	138	4	US-10-437-963-191916
9	449	22.1	120	4	US-10-767-701-57151
10	361	17.8	84	4	US-10-106-698-5464
11	324	16.0	162	4	US-10-424-599-196166
12	239	11.8	123	4	US-10-425-115-300949
13	221.5	10.9	107	4	US-10-425-115-230934
14	209	10.3	109	4	US-10-424-599-196168
15	149	7.3	278	4	US-10-425-114-46572
16	149	7.3	430	4	US-10-425-115-284883
17	142	7.0	381	4	US-10-437-963-194408
18	142	7.0	428	4	US-10-437-963-194408
19	134	6.6	375	5	US-10-739-930-6249
20	133.5	6.6	2052	4	US-10-437-963-137285
21	130.5	6.4	1212	5	US-10-788-793-2
22	127	6.3	1359	5	US-10-732-923-8707
23	127	6.3	1359	5	US-10-732-923-8708
24	126.5	6.2	785	4	US-10-425-115-332751
25	125.5	6.2	158	4	US-10-425-115-241450
26	124.5	6.1	1080	4	US-10-369-493-12560
27	124	6.1	726	5	US-10-483-680-2

ALIGNMENTS

RESULT 1

US-10-408-765A-1465
; Sequence 1465, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:

; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Boia D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Wartock, Dale E.

; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; IDENTIFIED IN THE MITOCHONDRIAL PROTEOME

; FILE REFERENCE: 660088.465

; CURRENT APPLICATION NUMBER: US/10/408,765A

; CURRENT FILING DATE: 2003-04-04

; NUMBER OF SEQ ID NOS: 3077

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1465

; LENGTH: 393

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-10-408-765A-1465

Query Match 93.4%; Score 1895; DB 4; Length 393;

Best Local Similarity 92.4%; Pred. No. 1.1e-137;

Matches 363; Conservative 14; Mismatches 14; Indels 2; Gaps 1;

Qy 1 MGKSDFLTPKAIANRIKSGKQLKRWYQCMQCRDENGFKCHMSHQRQLLASEN 60

Db 1 MGKSDFLTPKAIANRIKSGKQLKRWYQCMQCRDENGFKCHMSHQRQLLASEN 60

Qy 61 PQQFMDFSEEFNDLLELRRRFGTKRVHNNIYVNEYIISHREIHNNATQWETLTPTK 120

Db 61 PQQFMDFSEEFNDLLELRRRFGTKRVHNNIYVNEYIISHREIHNNATQWETLTPTK 120

Qy 121 WLGRGELCKVDPTPKGWYIYIDRDPETIRRLQLEKKKQDLDEEKTAKFIEQVRRG 180

Db 121 WLGRGELCKVDPTPKGWYIYIDRDPETIRRLQLEKKKQDLDEEKTAKFIEQVRRG 180

Qy 181 LEGKEQVPTFTLSRENBEKVTFTNLSKGACSSSGATSSKSTLGFPSALKTIGSSASVK 240

Db 181 LEGKEQVPTFTLSRENBEKVTFTNLSKGACSSSGATSSKSTLGFPSALKTIGSSASVK 240

Qy 241 RKSSQSSAQ--PAKKKKSALDEIMELEBEKKRTARTDAMLQPGIVVKVITTKLGEKYHK 298

Db 241 RKSSQSSQSSQ--PAKKKKSALDEIMELEBEKKRTARTDAMLQPGIVVKVITTKLGEKYHK 298

Qy 241 RKSSQSSQSSQ--PAKKKKSALDEIMELEBEKKRTARTDAMLQPGIVVKVITTKLGEKYHK 300

Db 241 RKSSQSSQSSQ--PAKKKKSALDEIMELEBEKKRTARTDAMLQPGIVVKVITTKLGEKYHK 300

Qy 299 KKGVVKEVIDRYTAVVKKMTSDGDLRLDQTHLETVIAPGKRVLVNLGGVYRGNGTLESI 358

Db 299 KKGVVKEVIDRYTAVVKKMTSDGDLRLDQTHLETVIAPGKRVLVNLGGVYRGNGTLESI 358

Qy 299 KKGVVKEVIDRYTAVVKKMTSDGDLRLDQTHLETVIAPGKRVLVNLGGVYRGNGTLESI 358

Db 299 KKGVVKEVIDRYTAVVKKMTSDGDLRLDQTHLETVIAPGKRVLVNLGGVYRGNGTLESI 358

Qy 299 KKGVVKEVIDRYTAVVKKMTSDGDLRLDQTHLETVIAPGKRVLVNLGGVYRGNGTLESI 358

Db 299 KKGVVKEVIDRYTAVVKKMTSDGDLRLDQTHLETVIAPGKRVLVNLGGVYRGNGTLESI 358

Db 301 KKAIVKEVDKTA VVVMDSGDKLDDQTHLETVIPAPGKRILVNLGNGYRGNEGTLSEI 360
QY 359 NEKAFSATIVETGPKLGRVREGIOYEDISKLA 391
Db 361 NEKTFSATIVETGPKLGRVREGIOYEDISKLA 393

RESULT 2

US-11-097-143-41040
; Sequence 41040, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 41040
; LENGTH: 390
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-41040

Query Match 51.9%; Score 1052.5; DB 6; Length 390;
Best Local Similarity 52.2%; Pred. No. 1.1e-72;
Matches 206; Conservative 69; Mismatches 107; Indels 13; Gaps 5;

QY 1 MGKSDFLSPKAIANRIKSKGLQKLRYWCQKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGRAEVGTPKYLANKMKSGLQKLRYWCQKQCRDENGFKCHCMSESHORQLLASEN 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATQMETLTDFTK 120
Db 61 PQKFLHSFKSEFSDGMEWLLRRRFGTKRVHNNIVYNEYISHREHIHMNATQMETLTDFTK 120
QY 121 WLREGCLKVDTPKGYIYIDRDPETIRROLELEKKKKQDLDDEEKTAKFIEQVR-- 178
Db 121 WLGRGTQVIADETEKGMFWYIDRSPAEKAKADKKEKMEKDDERWADFIEQIKN 180
QY 179 RLGEKQE--TPVFTLSR-ENEBEKKVTNLNKGAGGSAGATSKSSSLGSPALKLLGSA 236
Db 181 KAKDGEDEGQKFTBLKRENEPLKDIRLEK-----KFPDPTVLGKSALAKRPAP 232
QY 237 ASGRKESQSSQAQPAKKKKSALDEIMELEBEKK-RTARTDQWLQGIYVVKIITKLG 295
Db 233 EAEKVFKKPKSVAGDSQTRSVLDEIIKQBEESKKEKRNKRYWLHGIYVVKIITKSG 292
QY 296 YHKKGWVKEVDRYTAVVMTDSGDRLLKLDQTHLETVIPAGKRVLVNLGNGYRGNEGTL 355
Db 293 FFKQKAVLDVIDRYOGKIKFLETGKVKLDQAHLETVIPALDKPVMVNGAYRGEALL 352

QY 356 ESINEKAFSATIVETGPKLGRVREGIOYEDISKL 390
Db 353 RKLDERYSVSVEILLHGLPKGRIVDNVQYEDISKL 387

RESULT 3

US-10-425-114-48183
; Sequence 48183, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48183
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI.p
US-10-425-114-48183

Query Match 45.0%; Score 913; DB 4; Length 437;
Best Local Similarity 44.9%; Pred. No. 7.3e-62;
Matches 190; Conservative 75; Mismatches 126; Indels 32; Gaps 7;

QY 1 MGKSDFLSPKAIANRIKSKGLQKLRYWCQKQCRDENGFKCHCMSESHORQLLASEN 60
Db 14 MGKHEFLTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSESHORQVFGNA 73
QY 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATQMETLTDFTK 120
Db 74 PDRVVEGFSSEFLESFLSLIRRAHRSVAATVVYNEYIADRRHHVHMNSTWATLTFVK 133
QY 121 WLREGCLKVDTPKGYIYIDRDPETIRROLELEKKKKQDLDDEEKTAKFIEQVR- 179
Db 134 FLIGREGYCKVEDTPKGMFWYIDRDSQAVKRLKRIKSDMADDERQERMIARQIEKA 193
QY 180 -----GLEGEQETPVFTLSRENEBEKVTFNLNKGAGGSAG--ATTSSSSSLGP 227
Db 194 HKSLAKPNNGGNAAGEPESGEYSGSDDEQEPEDDSKADKATGKTAIALQKAVPGP 253
QY 228 SA-----LKL-----LGSAAAGKRK-----ESSQSSAQPAKKKKSALDEIMELEBE- 269
Db 254 KVNPPDDKPKMKFGPDEEDGSGTRDQEKNELTKKMGKDVKAAEAARSALDELMKEBEAK 313
QY 270 RTARTDQWLQGIYVVKIITKLGK- YHKKGWVKEVDIYTA VVKMTSDGRLKLDOT 328
Db 314 ESNRKKDQWLCFQGIYVVKVKSALAEKGYKQGVVKKVVDKYVGEIEMLESKHVLRVDD 373
QY 329 HLETVIPAGKRVLVNLGNGYRGNEGTLSEINEKAFSATIVETGPKLGRVREGIOYEDIS 388
Db 374 ELETVIPQIGGLVIRVNGAYRGSNARLLSVDTKEFCAKVQVEKGLYDGKVLRAVEYEDIC 433
QY 389 KLA 391
Db 434 KIS 436

RESULT 4

US-10-425-114-58982
; Sequence 58982, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong

APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 58982
LENGTH: 437
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: 700238628_FLI.pep
US-10-425-114-58982

Query Match 45.0%; Score 913; DB 4; Length 437;
Best Local Similarity 44.9%; Pred. No. 7.3e-62;
Matches 190; Conservative 75; Mismatches 126; Indels 32; Gaps 7;
Qy 1 MGKSDFLSPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 14 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFGMA 73
Qy 61 PQOFMDYFSEEFNRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATOWETLTDTFK 120
Db 74 PDVVEGFSEEFLESFLSLIRRAHRSRVAATVYNEYIADRHVHMNSTRWATLTFFVK 133
Qy 121 WLREGCLKVDTPKGYIYIDRDPTIRROLELEKXKQDLDDEKTAFTIEEQVRR- 179
Db 134 FLGREGYCKVEDTPKGMFTYIDRDSEQAVKRLKRIKSDMADDERQERMIARQIERA 193
Qy 180 -----GLEKQETPVFTLSRENEEKVTFLNKGAGGSAG--ATTSSSSSLGP 227
Db 194 HKSIAKPNGGGAEGEPESGSEYSGDDDEQEPEDSKADKATGKIAIALQKAVPGP 253
Qy 228 SA-----LKL-----LGSASAGRK-----ESSQSSAQPAKKKKSALDEIMELBEE-K 269
Db 254 KVNPFDDKPKMKFGDEEDSGTRDQEKNELTKMGKDVKAABAKRSALDELAKKEBEKAK 313
Qy 270 KRTARTDAMLQPGIVVVKIITKLGK- YHKKKGWVKEVIDRYTAVVQMTDSGRLKLDQT 328
Db 314 ERSNRKDYWLCPGIVVVKMSKSLAEKGYKQGVVKKVIDKYVGEIEMLESKHVLRVDQD 373
Qy 329 HLETVIPAGKRVLVNGYRGNEGTLSEINEKAFSATIVETGPKRRVVEGIQYEDIS 388
Db 374 ELETVPQIGGLVRIVNGAYRGSNARLLSVDTKEFCAKQVKEKGLYDGKVLRAVEYEDIC 433
Qy 389 KLA 391
Db 434 KIS 436

RESULT 5
US-10-425-115-341788
Sequence 341788, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 341788
LENGTH: 424

TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_74879C.1.pep
US-10-425-115-341788
Query Match 44.8%; Score 908; DB 4; Length 424;
Best Local Similarity 44.9%; Pred. No. 1.7e-61;
Matches 190; Conservative 73; Mismatches 128; Indels 32; Gaps 7;
Qy 1 MGKSDFLSPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFGMA 60
Qy 61 PQOFMDYFSEEFNRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATOWETLTDTFK 120
Db 61 PDVVEGFSEEFLESFLSLIRRAHRSRVAATVYNEYIADRHVHMNSTRWATLTFFVK 120
Qy 121 WLREGCLKVDTPKGYIYIDRDPTIRROLELEKXKQDLDDEKTAFTIEEQVRR- 179
Db 121 LLGREGYCKVEDTPKGMFTYIDRDSEQAVKRLKRIKSDMADDERQERMIARQIERA 180
Qy 180 -----GLEKQETPVFTLSRENEEKVTFLNKGAGGSAG--ATTSSSSSLGP 227
Db 181 HKSIAKPNGGGAEGEPESGSEYSGDDDEQEPEDSKADKATGKIAIALQKAVPGP 240
Qy 228 SA-----LKL-----LGSASAGRK-----ESSQSSAQPAKKKKSALDEIMELBEE-K 269
Db 241 KVNPFDDKPKMKFGDEEDSGTRDQEKNELTKMGKDVKAABAKRSALDELAKKEBEKAK 300
Qy 270 KRTARTDAMLQPGIVVVKIITKLGK- YHKKKGWVKEVIDRYTAVVQMTDSGRLKLDQT 328
Db 301 ERSNRKDYWLCPGIVVVKMSKSLAEKGYKQGVVKKVIDKYVGEIEMLESKHVLRVDQD 360
Qy 329 HLETVIPAGKRVLVNGYRGNEGTLSEINEKAFSATIVETGPKRRVVEGIQYEDIS 388
Db 361 ELETVPQIGGLVRIVNGAYRGSNARLLSVDTKEFCAKQVKEKGLYDGKVLRAVEYEDIC 420
Qy 389 KLA 391
Db 421 KIS 423
RESULT 6
US-10-739-930-8499
Sequence 8499, Application US/10739930
Publication No. US20040216190A1
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
FILE REFERENCE: 38-21(53377)B
CURRENT APPLICATION NUMBER: US/10/739,930
CURRENT FILING DATE: 2003-12-18
NUMBER OF SEQ ID NOS: 11088
SEQ ID NO 8499
LENGTH: 423
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C7583_2.p
US-10-739-930-8499
Query Match 44.5%; Score 901.5; DB 5; Length 423;
Best Local Similarity 44.5%; Pred. No. 5.4e-61;
Matches 189; Conservative 72; Mismatches 127; Indels 37; Gaps 8;
Qy 1 MGKSDFLSPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFGMA 60
Qy 61 PQOFMDYFSEEFNRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATOWETLTDTFK 120

```
Db 61 PDRVVEGSEFELESLIRRAHRSVAATVYVNEYIADRRHHVHMSTRWATLTFVK 120
QY 121 WLGREGLCVDETPKGYIYQIDRDETIRROLELEKKKKQDLDEEKTAKFIESQVRG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 LLGREGYCKVEDTPKGFMTYIDRSEQAVKDLKKRIKSDMAEDERQERMIARQIERA 180
QY 181 LE-----GKEQETPVFTELSENEE-EKYTFNLNKGAGGSAGA--TTS 220
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 HKSLAKPNDGDAAGEPESGGSEE---YSGDDGDEEDVSKAANKATGKIAIALQTAV 237
QY 221 KSSSLGPSALK-----LLGSAAGKRKE-----SSQSSAOPAKKKKALDEIMELER- 267
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 RQKNVPFDEKVEKFRPEEDGLTGGEKDELAKKKGDKVKAADARRSALDELKEEM 297
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 269 EKRTARTDAWIQPGIVKVIITKKLGEK-YHKKGGVVEVIDRYTAVVQMTSGDRLKLD 326
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 298 AKERSNRKDYWLCPGIVVKVMSKSLAEGYKQGVVRKMDKYVGEIEMLESKHVLRVD 357
QY 327 QTHLETVIPAPKRVVLVNGYRGNEGTLSEINEKAFSATIVIEGPKLGRRVEGIQYED 386
Db 358 QDELETVIPQIGGLVRILNGAYRGSNARLLSVDTKEFCAKVQVERGLYDQVLRVAVKYBD 417
QY 387 ISKLA 391
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 418 ICKIS 422

RESULT 7
US-10-424-599-168326
; Sequence 168326, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 168326
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1.pep
US-10-424-599-168326

Query Match 30.2%; Score 613; DB 4; Length 176;
Best Local Similarity 64.3%; Pred. No. 3.1e-39;
Matches 108; Conservative 26; Mismatches 34; Indels 0; Gaps 0;

QY 1 MGKSDFLSPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHORQLLASEN 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MGKNEFLTPTKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSEGHQRMQIFQGN 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGKRVHNNIVVNEYISHREHIHMNATOWETLTDFTK 120
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 PHRIVEGYSEEPSTFLEHMKSHRSRVAATVYVNEYINDRRHHIMNSTQWATLTFVK 120
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 WLGREGLCVDETPKGYIYQIDRDETIRROLELEKKKKQDLDEEK 168
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 YLGRTGCKVEETPKGFMTYIDRSETLTKERMKNKRIKADWDVEEK 168

RESULT 8
US-10-437-963-191916
; Sequence 191916, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

```
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 191916
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1.pep
US-10-437-963-191916

Query Match 26.2%; Score 531; DB 4; Length 138;
Best Local Similarity 66.7%; Pred. No. 4.7e-33;
Matches 92; Conservative 22; Mismatches 24; Indels 0; Gaps 0;

QY 1 MGKSDFLSPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHORQLLASEN 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MGKHEFLTPTKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSESHORQMVFQQA 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGKRVHNNIVVNEYISHREHIHMNATOWETLTDFTK 120
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 PDRVVEGSEFEFLDAFLTLRRRAHRSIAATVYVNEFIADRRHHVHMSTRWATLTFVK 120
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 WLGREGLCVDETPKGY 138
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 FLGREGHCKVEDTPKGNF 138

RESULT 9
US-10-767-701-57151
; Sequence 57151, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 57151
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30963818.pep
US-10-767-701-57151

Query Match 22.1%; Score 449; DB 4; Length 120;
Best Local Similarity 66.7%; Pred. No. 8.4e-27;
Matches 80; Conservative 15; Mismatches 25; Indels 0; Gaps 0;

QY 1 MGKSDFLSPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHORQLLASEN 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 MGKHEFLTPTKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSESHORQMVFQGA 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGKRVHNNIVVNEYISHREHIHMNATOWETLTDFTK 120
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 PDRVVEGSEFEFLSLIRRAHRSVAATVYVNEYIADRRHHVHMSTRWATLTFVK 120
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 10

US-10-106-698-5464
; Sequence 5464, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005PI
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5464
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5464

Query Match 17.8%; Score 361; DB 4; Length 84;
Best Local Similarity 94.6%; Pred. No. 3.2e-20;
Matches 70; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 318 DSGDKLKDQTHLETVPAPCKRVLVNGGYRGNEGTLSEINEKAFSATIVETGPLKGR 377
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 11 ESGDKLKLDQTHLETVPAPGKRILVNGGYRGNEGTLSEINEKTFSATIVETGPLKGR 70
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

Qy 378 RVEGIQYEDISKLA 391
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 71 RVEGIQYEDISKLA 84
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

RESULT 11

US-10-424-599-196166
; Sequence 196166, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196166
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19163C.1.pep
US-10-424-599-196166

Query Match 16.0%; Score 324; DB 4; Length 162;
Best Local Similarity 46.8%; Pred. No. 5.5e-17;
Matches 72; Conservative 27; Mismatches 45; Indels 10; Gaps 3;

Qy 240 KRKSSQSQAQPAKKKSALDEIMELEBEKK-RTARTDAWLQPGIIVVKIITKLGEK-YH 297
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 17 KRKESGGG-----KSALEDMMREBEKKENRKDYWLHEGIVVVKVMKSLAEKGYY 68
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

Qy 298 KKKGWKEVIDRYTAVVWKMTDSGRLKLDQTHLETVIPAPCKRVLVNGGYRGNEGTLSE 357
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 69 KOKGVVRKIDKYGEIEMLSESKHVLRVDQALETVIPQVGRRKVINGATRGSIALLG 128
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

Qy 358 INEKAFSATIVETGPLKGRVEGIQYEDISKLA 391
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:28:34 ; Search time 4.98725 Seconds
(without alignments)
237.628 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 2028
Sequence: 1 MGKSDFLSPKAIANRIKSG.....GPKGRVRGVIQVEDISKLA 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 17545 seqs, 3030971 residues

Total number of hits satisfying chosen parameters: 17545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New:
1: /cgn2_6/ptodata/2/pubpa/US10_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpa/US06_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpa/US07_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpa/US08_NEW_PUB.pep.*
5: /cgn2_6/ptodata/2/pubpa/US09_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpa/PCT_NEW_PUB.pep.*
7: /cgn2_6/ptodata/2/pubpa/US11_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	110.5	5.4	853	1	US-10-821-234-1110
2	105	5.2	885	1	US-10-793-626-1660
3	104	5.1	676	7	US-11-135-855-28
4	104	5.1	709	7	US-11-074-176-158
5	104	5.1	717	7	US-11-135-855-29
6	100	4.9	672	1	US-10-689-742-70
7	99.5	4.9	1448	1	US-10-485-517-212
8	94.5	4.7	583	1	US-10-793-626-1358
9	93.5	4.6	756	7	US-11-074-176-202
10	93.5	4.6	1189	7	US-11-074-176-134
11	93.5	4.6	1279	1	US-10-793-626-3188
12	93	4.6	989	1	US-10-821-234-975
13	92.5	4.6	752	1	US-10-793-626-348
14	91	4.5	677	1	US-10-982-545-12
15	90.5	4.5	618	1	US-10-793-626-860
16	89.5	4.4	422	1	US-10-821-234-1313
17	88.5	4.4	296	1	US-10-131-826A-26
18	87.5	4.3	1618	1	US-10-984-645-2
19	87	4.3	416	1	US-10-793-626-2
20	87	4.3	1437	7	US-11-074-176-96
21	87	4.3	5024	1	US-10-793-626-2964
22	86.5	4.3	594	1	US-10-131-826A-10
23	86.5	4.3	594	1	US-10-485-517-416
24	86.5	4.3	1299	1	US-10-821-234-1145
25	86.5	4.3	2897	1	US-10-499-715-2

26	85	4.2	419	1	US-10-793-626-2516	Sequence 2516, Ap
27	85	4.2	745	1	US-10-793-626-1500	Sequence 1500, Ap
28	84.5	4.2	309	7	US-11-074-176-276	Sequence 276, App
29	84	4.1	431	7	US-11-074-176-132	Sequence 132, App
30	84	4.1	770	1	US-10-982-545-15	Sequence 15, Appl
31	84	4.1	770	1	US-10-789-273-38	Sequence 38, Appl
32	83.5	4.1	654	7	US-11-046-668-4	Sequence 4, Appl
33	83.5	4.1	1614	1	US-10-821-234-903	Sequence 903, App
34	83	4.1	930	1	US-10-821-234-1188	Sequence 1188, Ap
35	82.5	4.1	691	1	US-10-131-826A-16	Sequence 16, Appl
36	82.5	4.1	703	1	US-10-821-234-963	Sequence 963, App
37	81.5	4.0	231	1	US-10-957-569-21	Sequence 21, Appl
38	81	4.0	196	1	US-10-793-626-2088	Sequence 2088, Ap
39	80.5	4.0	752	1	US-10-793-626-1138	Sequence 1138, Ap
40	80.5	4.0	874	7	US-11-012-762-8	Sequence 8, Appl
41	80.5	4.0	974	1	US-10-821-234-1152	Sequence 1152, Ap
42	80	3.9	469	1	US-10-821-234-1151	Sequence 1151, Ap
43	80	3.9	615	1	US-10-982-545-14	Sequence 14, Appl
44	80	3.9	674	1	US-10-507-275-9	Sequence 9, Appl
45	80	3.9	683	7	US-11-046-668-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-10-821-234-1110
; Sequence 1110, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1110
; LENGTH: 853
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-821-234-1110

Query Match	5.4%	Score 110.5;	DB 1;	Length 853;
Best Local Similarity	28.3%	Pred. No. 0.26;		
Matches	58;	Conservative	24;	Mismatches 58; Indels 65; Gaps 11;
Qy	120	KWLGREGLC-KVDETPKGVWYQIDRDPETIRQLELEKK-----KK	160	
Db	561	KALWFQRCVKVLSBK--YKLVLRIP---NRGIDLLKKDKRKRYSYSPDGKESPSDKK	615	
Qy	161	QDLDEDEKTAFFI---EEQVRGLEK-----EQTPTFTLSRE---NEBEKVTFN	206	
Db	616	SKTDGSKQTESSTEGKEQEKSGEDGKOTKDDQTEQEPNNLLESDELLVDEEE	670	
Qy	207	LNKGAGGSAGATTSKSSLGPSA-LKLGSAAASKRKESQ-----SSAQPAKKKKA	258	
Db	671	-----AALLLESGSVGDETDLANLGDVADGKKEPSDKAVKKGDSAAAKKLLK	722	
Qy	259	LDEIMELEEKKTARTDAWLQPGI	283	
Db	723	VDKIEELDQB-----NEAALengi	741	

RESULT 2
US-10-793-626-1660
; Sequence 1660, Application US/10793626
; Publication No. US20050255478A1

Db 145 DMVAGSLRGSFPERLKKVIDKAKSDPNIVLFDLHNIVGAGSTDSENN-GDAANILK 203
Qy 234 GSAAGKRKSSOSSAQAQPAKKKXALDEIMELBEEKRTARTDANLQ-----GIVVKII- 288
Db 204 PALASGLKIJGATTT-----SEFORIEKDPALSRFQAVQVPEPSTDAIKILE 253
Qy 289 -TKKLGKHYHKKGVVKEVIDRYTAVVKMTSDGRKLQDTHLETVIPA---PGKRVLV 344
Db 254 GLKKYEDYH-----VKYTD--DSLKLAVELSERVIOGRYLPDKAIDLM 296
Qy 345 NGYRGNEGTLSEINAKAFSATIVETGPLKGRRVEGIIQYEDISKLA 391
Db 297 DEAGAKKALLVQPTDEKSLKNQI-----SALEAKKAAEAAKAEVDYDKA 339

RESULT 5

US-11-135-855-29
; Sequence 29, Application US/11135855
; Publication No. US2005025557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-29

Query Match 5.1%; Score 104; DB 7; Length 717;
Best Local Similarity 21.3%; Pred. No. 0.63;
Matches 61; Conservative 46; Mismatches 89; Indels 90; Gaps 12;
Qy 144 RDPETIRROLEKKKKQDLDEEKTAKFIEEQVRRGLEKGEQETPVFTLSRENEBEKV 203
Db 359 RRDARRRLRARRREQ-----EEELRLRLEQKEKEKERRERADRGAE-- 404
Qy 204 TFLNKGAGGAGATTSSKSLGSPALKLGSAAAGKRKSSQAQPAKKKXALDEIM 263
Db 405 -----RSGSGSGDELREDD---EPVKRGRKGRGPPSSSDSEPEA----- 444
Qy 264 ELEBEKGTAR-----TDWLQPGIVVKII-----TKKLGKY----- 296
Db 445 ELEREAKKSAKPOSSSTEPARKGQKEKVRPEEKQAQPKVKTTRKRGSEGSMDRKV 504
Qy 297 -HKKGGVKEVIDRYTAVVKMTSDGRKLQDTHLETVIPA-----PGKRVLVINGG 347
Db 505 EKKKEPSVEEKLQKLSIKFA-----LKVSPDVKRCINALEELGTQVTSQILQKNTD 559
Qy 348 -----YRGNEGTLSEINAKAFSATIVETGPLKGRRVEGIIQ 383
Db 560 VVATLKKIRRYKAN---KDVMEKAAEVYTRLKSRLV-GPKIEAVQ 600

RESULT 6

US-10-689-742-70
; Sequence 70, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: LaVallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766-000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 70
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (107)..(107)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (111)..(111)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117)..(118)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (645)..(645)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-10-689-742-70
Query Match 4.9%; Score 100; DB 1; Length 672;
Best Local Similarity 23.6%; Pred. No. 1.1;
Matches 70; Conservative 41; Mismatches 107; Indels 78; Gaps 14;
Qy 33 KQCRDENGFKCHCMSESHORQLLASENPOQFMDYFSEEPNDFLELLRRRFGTKRVHNN 92
Db 424 QQCAEEEG-----DIEWQLNASVLVAQAQVKLRDF--ESAVNNFEKALER---AKLVHNN 473
Qy 93 IVNEYISHREHIHMNATQWETLTDFTYKWLGRGLCKVD-----ETPKGWYIOYIDR-- 144
Db 474 EAQAQIIS-----ALDDANKGIIRE-LRKNTYVENLKEKSEGEASLYEDRII 519
Qy 145 -----DPTIRROLE-LEKKKKQDLDEEKTAKFIEEQVRRGLEKGEQETPVFT 192
Db 520 TREKDMRVDRDEPKVQKQMDHSEDEKETDEDEA-----FGEALQSPAGSKQSV----- 569
Qy 193 ELSRENEEBEKVTFNLNKGAGGAGATTSSKSLGSPALKLGSAAAGKRKSSQAQPA 252
Db 570 -----EAGKARSDLGAVAKGLSGELGTRS---GETGRKLLLEAG---RRSREIYRPS 616
Qy 253 KKKKKSAL-----DEIMELBEEKRTARTDANLQPGIVVKIIITKKLGKHYHK 299
Db 617 GELFQRLSGBFSROPEPELKKLSEVGRRXPEELGKTQFGEIGE--TKKTGNEMSEKE 670

RESULT 7

US-10-485-517-212
; Sequence 212, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides

Best Local Similarity 17.4%; Pred. No. 6.2;		Matches 75; Conservative 84; Mismatches 145; Indels 128; Gaps 16;	
QY	49	SHORQLLASENPOQPMY-----FSEFRNDLELLRRFGTKRVHNNI	93
Db	510	SVQKMKGLVSPESMDNYSHPHELAVTBEEINVLDQDLQNALSESRNEKVELEEK	569
QY	94	VYNE-----YISHREHHMNAQTWETLDTFKWL-GREGLCCKVDETP	134
Db	570	VEREKGTVIKPPVEEYEMKSSYCVIENWK-----EKAFLPEKYQEAQBEIMKLKDTL	624
QY	135	KGYIYQIDRDPETIRQL-----ELEKKKKQDLDP-----SEKATXP	172
Db	625	KSQMTQASDEADMKEMAMNMIDELNKQVSELSQLYKEAQAELEDRKRKSLSDVTARY	684
QY	173	IEQVRRGL-----EGKEQETVPFTEL-----SRENEEBKVTFLNLN	208
Db	685	IHKAEHEKLMQNTVNSRAKEDALSEMKSQYSKVLNLTQLQLVDAQKENSVSITEHLQ	744
QY	209	KGAGGSAGATTSSSLGSPALKLGLSAAAGKRGKSSQSAQPAKKKXGALDEIMLEBEE	268
Db	745	V-----ITTLR-----TAAKEMBEKISNLKHLASKEVEVAKLEKQLLEE-----	784
QY	269	KKTARTDAW-----LQPGI--VVKIITKLGKHYHKKGVKVEVDIYTAUVKMTD	318
Db	785	--KAAMTDAMVPRSSYEKLQSSLESEVSLASKLESVKEKEV-----HSEVVQIRS	835
QY	319	SGDRKLDOHLETVIPAPKRVLVNGGYRGNEGTLSEINAKAFSAFTIVETGPKLGRR	378
Db	836	EVSQVREKENIOTLKSKEQEVNELLQKFOQAQBELAEMKRYAESSSKLEED---KDKK	892
QY	379	VEGIOVEDISKL 390	
Db	893	INEMS-KEVTKL 903	
RESULT 13			
US-10-793-626-348			
; Sequence 348, Application US/10793626			
; Publication No. US20050255478A1			
; GENERAL INFORMATION:			
; APPLICANT: KIMMERLY, WILLIAM JOHN			
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS			
; FILE REFERENCE: PU3480US			
; CURRENT APPLICATION NUMBER: US/10/793,626			
; CURRENT FILING DATE: 2004-03-04			
; PRIOR APPLICATION NUMBER: 60/164,258			
; PRIOR FILING DATE: 1999-11-09			
; NUMBER OF SEQ ID NOS: 4472			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 348			
; LENGTH: 752			
; TYPE: PRT			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: synthetic			
; OTHER INFORMATION: amino acid sequence			
US-10-793-626-348			
Query Match		4.6%;	Score 92.5; DB 1; Length 752;
Best Local Similarity		18.1%;	Pred. No. 4.7;
Matches		62; Conservative	73; Mismatches 112; Indels 95; Gaps 16;
QY	123	GREGLCVKDETPKGYI--QYIDRDP-ETIRRL-----ELEK-----KKQDLDDDEEK	168
Db	174	GIEGIKAVETGKGVVRSRVDEEPLRSRKGKLIITEIPYEVNKSLSLVKRIDELRADKK	233
QY	169	TAKFIE---EQVRGLE-----QKEQETVPFTELSRENEEBKVTFLNKGAGGSAGATTS	220
Db	234	VDGIVEVRDETDRTGLRIATIELKDNANSEIKNLYKNSDLQISYFNFMVA-----	284
QY	221	KSSSLGFSALKLIG-----SASAGRKESQSSQAQPAKKKKSALDE	261
Query Match		4.5%;	Score 91; DB 1; Length 677;
Best Local Similarity		22.6%;	Pred. No. 5.3;

Db	205	-----ISEGRPKLMGLREILIESYLNHQIEVVTNRTRYDLEQAEKRMHIIVEGLMKALSILDE	340
QY	262	IMEL--EEBKRTAR-----TDAWLQPGIIVVKII-----TKLGEKYHKKGVVK 304	
Db	341	VIALIRNSKNKKDAKDNLVAEYDFTFEAQAEALVMLQLVRLTWTDTIEALKKEHEELEALIK	400
QY	305	E---VIDRYTAV-----VKMTDSGDRL-----KLDQTHLETVIPAPKRVLVING	346
Db	401	ELRNLDNHEALLAVIKDELNEIKKFKVDRLSTTIEAIEISEIKDKVMVPSEVIL---	457
QY	347	GYRGNEGTLSEINAKAFSAFTIVETGPKLGRRVEGIOVEDIS	388
Db	458	SLTOHQYIKRTSTRSFNASGVTEIGLKQDGRL---LKHSVN	496
RESULT 14			
US-10-982-545-12			
; Sequence 12, Application US/10982545			
; Publication No. US20050244890A1			
; GENERAL INFORMATION:			
; APPLICANT: Davies, Huw Alun			
; APPLICANT: McGuire, James			
; APPLICANT: Simonsen, Anja Hviid			
; APPLICANT: Blennow, Kaj			
; APPLICANT: Podust, Vladimir			
; APPLICANT: Ciphergen Biosystems, Inc.			
; TITLE OF INVENTION: Biomarkers for Alzheimer's Disease			
; FILE REFERENCE: 016866-011550US			
; CURRENT APPLICATION NUMBER: US/10/982,545			
; CURRENT FILING DATE: 2004-11-06			
; PRIOR APPLICATION NUMBER: US 60/518,360			
; PRIOR FILING DATE: 2003-11-07			
; PRIOR APPLICATION NUMBER: US 60/526,753			
; PRIOR FILING DATE: 2003-12-02			
; PRIOR APPLICATION NUMBER: US 60/546,423			
; PRIOR FILING DATE: 2004-02-19			
; PRIOR APPLICATION NUMBER: US 60/547,250			
; PRIOR FILING DATE: 2004-02-23			
; PRIOR APPLICATION NUMBER: US 60/558,896			
; PRIOR FILING DATE: 2004-04-02			
; PRIOR APPLICATION NUMBER: US 60/572,617			
; PRIOR FILING DATE: 2004-05-18			
; PRIOR APPLICATION NUMBER: US 60/586,503			
; PRIOR FILING DATE: 2004-07-08			
; NUMBER OF SEQ ID NOS: 16			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 12			
; LENGTH: 677			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
; FEATURE:			
; OTHER INFORMATION: Chromogranin B precursor			
; FEATURE:			
; NAME/KEY: SIGNAL			
; LOCATION: (1)..(20)			
; OTHER INFORMATION: signal peptide			
; FEATURE:			
; NAME/KEY: PEPTIDE			
; LOCATION: (21)..(677)			
; OTHER INFORMATION: Chromogranin B (Secretogranin I)			
; FEATURE:			
; NAME/KEY: PEPTIDE			
; LOCATION: (326)..(385)			
; OTHER INFORMATION: biomarker peptide 7258 Da, processed fragment of			
; OTHER INFORMATION: Chromogranin B			
; FEATURE:			
; NAME/KEY: MOD RES			
; LOCATION: (341)			
; OTHER INFORMATION: Xaa = sulfo tyrosine			
US-10-982-545-12			

Matches	60;	Conservative	33;	Mismatches	119;	Indels	54;	Gaps	10;
Qy	111	QWETLTDTKWLGRGLCKVDETPKGYIQYIDRDPETIRROLELEKKKKQDLDEEKA	170						
Db	113	QGPTKADTEKWAEGGHSRERADFPQWSLPDSQVSEVVKTRHSEKSDQDEDEEEE---	169						
Qy	171	KFIEEQVRRGLLEG-----KEQETPVTFELSGRE-----EKKVTFNLNGAGGSA	215						
Db	170	---GENYQKGERGEDSSEKHLEFPGTQNAFLNERKQASAIKKEELVARETHAAGHSQ	226						
Qy	216	GATTSKSSSLGPSALKLUGSAAQGRKXESSQSAQPAKKKKSAL-DEIMLELEEKRT--	272						
Db	227	EKTHSRKSSQES-----GEEAGSQENHPQESKGQPRSQSESEEGEEDATSEVDKRTRP	281						
Qy	273	-----ARTDAWLPGIIVVKLIITKLGKYHKKGKGVKVEDRYTAVVYKMTDSGDRLKLD	326						
Db	282	RHHGHRGRDRSSQGSILPS-----EKKHPQE--ESESNSVMSASLGEKRDHH	328						
Qy	327	QTHLETVIPAF--GKRVLVLLNGGYRG	350						
Db	329	STHYRASEEEPEXGEEI-----KGYPG	350						

RESULT 15

```

US-10-793-626-860
; Sequence 860, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 860
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-860

```

Query Match	4.5%;	Score 90.5;	DB 1;	Length 618;
Best Local Similarity	18.5%;	Pred. No. 5.1;		
Matches	70;	Conservative 50;	Mismatches 130;	Indels 129; Gaps 16;
QY	3	KSDFLSPKAIARIRKSKGLQKLRYWCQCKQCRDENGFKCHCMSESHOROLLASE---	59	
DB	275	KNVVDPNVLIDRY---GLDATTRYL-----MRFLPGSDGVF	309	
QY	60	NPOQFMDYFSEBFRNDFLELLRRRFG--TKRVHNNIVTNEYISHREHIHMNATQMTLTD	117	
DB	310	TPEAFVERTNYDLANDLGNLVNRTISMINKYFHGELPAYQGPKHDELDEKMEAMALETVKS	369	
QY	118	FT-----KWLGRGLCKVDETPKGWYI-----	139	
DB	370	FNDNMESLQFVSALSTVWKFISRTNK-YIDET-QPWVLAKDENOREMLGNVMAHLVENIR	427	
QY	140	-----QYIDRDPTIRROLELEKKKKQDLDEKTAKFTEEQVRRGLEGEQETPVFT	192	
DB	428	FATILLQLPFLTHAPREIPKQLNNPNDLHQLDLSLQQYGMLSB-----AITVTEKPTPIFP	482	
QY	193	ELSRNENBEKVTFNLNKAGGSAGATTSKSSSLGPSALKLLGSSAASGKR-----	241	
DB	493	RLDTAEATAYI-----KESMQPPKSIK--QSDPEPGKEQIDIKDFDKVE	523	
QY	242	-KSSQSSQAQPAKKKKSALDEIMELBEEKRTARTDA-WLQP-----GIWVKIIT-----K	290	
DB	524	IKAAATIIDAEENVYKSEKLLKIKVELDNQSRQIVSGIAKFPYRPEDITIGKVAVWVNLKPAAK	583	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:33:44 ; Search time 132.661 Seconds
(without alignments)
1295.011 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 391
Sequence: 1 MGKSDFLSPKAIANRIKSG.....GPKLGRVVGIOVEDISKLA 391

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 2443163 seqs, 439378781 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq 21:*
1: Genesexp1980s:*
2: Genesexp1990s:*
3: Genesexp2000s:*
4: Genesexp2001s:*
5: Genesexp2002s:*
6: Genesexp2003as:*
7: Genesexp2003bs:*
8: Genesexp2004s:*
9: Genesexp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	371	94.9	391	2	AAY23619 Murine ki
2	179	45.8	393	2	AAY23620 A human k
3	179	45.8	393	7	ADJ69659 Human hea
4	163	41.7	291	2	AAY23617 Mouse tru
5	120	30.7	293	2	AAY23618 A huamn t
6	42	10.7	390	8	ADP22452 Sea-squir
7	35	9.0	411	3	AG42576 Arabidops
8	35	9.0	423	3	AAB03064 Maize KIN
9	35	9.0	423	3	AAB03065 Maize KIN
10	35	9.0	423	8	ADT58422 Plant pol
11	35	9.0	424	3	AAB03063
12	35	9.0	437	8	ADX96318 Plant ful
13	35	9.0	437	8	ADX78817 Plant ful
14	29	7.4	84	4	AG74690 Human col
15	24	6.1	382	3	AG42577 Arabidops
16	21	5.4	302	3	AG46712 Arabidops
17	21	5.4	340	3	AG46711 Arabidops
18	21	5.4	343	3	AG46710 Arabidops
19	18	4.6	34	5	ADK35903 Novel hum
20	18	4.6	38	5	ADK35871 Novel hum
21	15	3.8	15	2	AAR66765 RecA-like
22	15	3.8	390	4	ABB71416 Drosophil
23	11	2.8	11	2	AAR66766 Zinc-fing
24	9	2.3	535	8	ADT60584 Plant pol

25	9	2.3	1542	5	ABE78013	Abb78013 Amino aci
26	9	2.3	3141	7	ADJ70444	Adj70444 Human hea
27	9	2.3	3144	2	AAR58777	Aar58777 Protein e
28	9	2.3	3144	2	AAW36887	AAW36887 Previousl
29	9	2.3	3144	2	AAW09871	AAW09871 Human hun
30	9	2.3	3144	2	AAW44742	AAW44742 Human hun
31	9	2.3	3144	2	AAY33493	Aay33493 Human hun
32	9	2.3	3144	9	ADY98141	Ady98141 Human hun
33	9	2.3	3223	4	ABB11407	Abb11407 Human Hun
34	9	2.3	3223	4	ABB11470	Abb11470 Human Hun
35	8	2.0	8	5	ABG96595	Abg96595 Human leu
36	8	2.0	8	5	ABG97218	Abg97218 Human leu
37	8	2.0	8	7	ADL99055	Adl99055 Human leu
38	8	2.0	8	7	ADL98437	Adl98437 Human com
39	8	2.0	10	4	AAG97042	Aag97042 Human com
40	8	2.0	159	3	AAG19456	Aag19456 Arabidops
41	8	2.0	173	4	AAU58581	Aau58581 Propionib
42	8	2.0	173	6	ABM55100	Abm55100 Propionib
43	8	2.0	179	8	ADJ93930	Adj93930 Human mit
44	8	2.0	183	4	ABB66749	Abb66749 Drosophil
45	8	2.0	192	4	ABB58711	Abb58711 Drosophil

ALIGNMENTS

RESULT 1
AAY23619
ID AAY23619 standard; protein; 391 AA.
XX
AC AAY23619;
DT 07-SEP-1999 (first entry)
XX
DE Murine kin17 protein.
XX
KM Mouse; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme.
XX
OS Mus sp.
XX
PN FR2772046-AL.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
XX
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
XX
Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
WPI; 1999-359999/31.
XX
New DNA coding for human kin17 protein - useful for controlling cell
proliferation or fertility.
PT
Claim 19; Page 36-37; 69pp; French.
XX
The present sequence represents a murine kin17 protein. The mammalian
kin17 protein is useful for preparing a medicament for controlling cell
proliferation or for controlling fertility. The medicaments can also be
used to treat hyperproliferative diseases. Fragments between amino acids
55 and 235 (preferably between amino acids 129 and 228) of a mammalian
kin17 protein are useful for regulating the interaction between proteins
and curved DNA. The fragment can be used to block replication of HIV or
its integration into the human genome or to target repair enzymes to
curved DNA sites. Expression vectors for kin17 can be used for
controlling cell proliferation
XX
Sequence 391 AA;

Query Match	94.9%;	Score 371;	DB 2;	Length 391;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 371;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	21	LQKLRYTCMCQCKQCRDENGFKCHCMSSHQHQLLASSENQQQFMDYFSEPRNDFLELL	80
Db	21	LQKLRYTCMCQCKQCRDENGFKCHCMSSHQHQLLASSENQQQFMDYFSEPRNDFLELL	80
Qy	81	RRRFGTKRVHNNIVYNEVISHREHIMHNAQTWETLTDFTKWLGRGLCKVDETPKGWYIQ	140
Db	81	RRRFGTKRVHNNIVYNEVISHREHIMHNAQTWETLTDFTKWLGRGLCKVDETPKGWYIQ	140
Qy	141	YIDRDPETIRRQLELEKKKKQDLDDDEKTAKFIEEQVRRGLEGKEQETPVFTELSRENEE	200
Db	141	YIDRDPETIRRQLELEKKKKQDLDDDEKTAKFIEEQVRRGLEGKEQETPVFTELSRENEE	200
Qy	201	EKVTFNLKAGGAGGAGATTSSKSSIGPSALKLGSAASGKKKSESSQSAQPAKKKKSALD	260
Db	201	EKVTFNLKAGGAGGAGATTSSKSSIGPSALKLGSAASGKKKSESSQSAQPAKKKKSALD	260
Qy	261	EIMELESEKKKTARTDAWLQPGIVVKKIITKKLGEKYHKKKGVVKEVIDRYTAVVKTDSG	320
Db	261	EIMELESEKKKTARTDAWLQPGIVVKKIITKKLGEKYHKKKGVVKEVIDRYTAVVKTDSG	320
Qy	321	DRKLQDTHLETVIPAPGKRVLVNLGGYRGNEGTLSEINAKFSATIVIEGTPLKGRRVE	380
Db	321	DRKLQDTHLETVIPAPGKRVLVNLGGYRGNEGTLSEINAKFSATIVIEGTPLKGRRVE	380
Qy	381	GIQVEDISKLA 391	
Db	381	GIQVEDISKLA 391	

RESULT 2	
AAAY23620	ID AAY23620 standard; protein; 393 AA.
XX	AAAY23620;
XX	07-SEP-1999 (first entry)
XX	A human kin17 protein.
XX	Human; kin17 protein; cell proliferation; fertility;
KW	hyperproliferative disease; protein interaction; curved DNA;
KW	HIV replication; HIV integration; repair enzyme.
XX	
OS	Homo sapiens.
XX	
XX	FR2772046-A1.
PN	
XX	11-JUN-1999.
PD	
XX	09-DEC-1997; 97FR-00015536.
PF	
XX	09-DEC-1997; 97FR-00015536.
PR	
XX	(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
PA	
XX	Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
PI	
XX	WPI; 1999-359999/31.
XX	
DR	
XX	New DNA coding for human kin17 protein - useful for controlling cell
PT	proliferation or fertility.
XX	
PS	Claim 19; Page 37-38; 69pp; French.
XX	

CC The present sequence represents a human kin17 protein. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids

CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, neurotropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteoparathic, ophthalmological and
 CC cyostatic activities. This polypeptide sequence is a human heart
 CC mitochondrial protein of the invention.

XX Sequence 393 AA;
 SQ
 Query Match 45.0%; Score 179; DB 7; Length 393;
 Best Local Similarity 100.0%; Pred. No. 4.5e-164;
 Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 PKAIANRIKSGLOKLRYWYCOMQKQCRDENGFKCHMSHQRLLASENPQQFMDYF 68
 Db 9 PKAIANRIKSGLOKLRYWYCOMQKQCRDENGFKCHMSHQRLLASENPQQFMDYF 68

Qy 69 SEBFRNDFLELRRTFKRVHNNIVYNEVISHREIHMNATOWETLDTFTKWLREGLC 128
 Db 69 SEBFRNDFLELRRTFKRVHNNIVYNEVISHREIHMNATOWETLDTFTKWLREGLC 128

Qy 129 KVDETPKGWYIQYIDRDPETIRRLQLELEKKKKQDLDEEKTAKFIEBQVRRGLEKEQE 187
 Db 129 KVDETPKGWYIQYIDRDPETIRRLQLELEKKKKQDLDEEKTAKFIEBQVRRGLEKEQE 187

RESULT 4
 AAY23617
 ID AAY23617 standard; protein; 291 AA.
 AC AAY23617;
 XX 07-SEP-1999 (first entry)
 XX Mouse truncated kinl7 protein.
 XX kinl7 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme.

OS Mus sp.
 XX FR2772046-A1.
 XX 11-JUN-1999.
 XX 09-DEC-1997; 97FR-00015536.
 XX 09-DEC-1997; 97FR-00015536.
 XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 XX WPI; 1999-359999/31.
 XX New DNA coding for human kinl7 protein - useful for controlling cell
 XX proliferation or fertility.
 XX Claim 14; Page 33-34; 69pp; French.

XX The present sequence represents a mouse kinl7 protein with amino acids
 CC 129-228 deleted. The mammalian kinl7 protein is useful for preparing a
 CC medicament for controlling cell proliferation or for controlling
 CC fertility. The medicaments can also be used to treat hyperproliferative
 CC diseases. Fragments between amino acids 55 and 235 (preferably between

CC amino acids 129 and 228) of a mammalian kinl7 protein are useful for
 CC regulating the interaction between proteins and curved DNA. The fragment
 CC can be used to block replication of HIV or its integration into the human
 CC genome or to target repair enzymes to curved DNA sites. Expression
 CC vectors for kinl7 can be used for controlling cell proliferation

XX Sequence 291 AA;
 SQ
 Query Match 41.7%; Score 163; DB 2; Length 291;
 Best Local Similarity 100.0%; Pred. No. 1e-148;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 229 ALKLLGSAASGRKSSQSSAQPAPKKKGALDEIMELEBEKKRTARTDAWLQPGIVVKII 288
 Db 129 ALKLLGSAASGRKSSQSSAQPAPKKKGALDEIMELEBEKKRTARTDAWLQPGIVVKII 188

Qy 289 TKKLGEKHYHKKGVVKEVIDRYTAVVYKMTDSGDRKLQDTHLETVPAPGKRVLVNGGY 348
 Db 189 TKKLGEKHYHKKGVVKEVIDRYTAVVYKMTDSGDRKLQDTHLETVPAPGKRVLVNGGY 248

Qy 349 RNEGTLSEINEKAFSATIVETGPKGRRVGGIOVEDISKLA 391
 Db 249 RNEGTLSEINEKAFSATIVETGPKGRRVGGIOVEDISKLA 291

RESULT 5
 AAY23618
 ID AAY23618 standard; protein; 293 AA.

AC AAY23618;
 XX 07-SEP-1999 (first entry)

XX A huamn truncated kinl7 protein.

XX kinl7 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme.

OS Homo sapiens.
 XX FR2772046-A1.
 XX 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.
 XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 XX WPI; 1999-359999/31.

XX New DNA coding for human kinl7 protein - useful for controlling cell
 XX proliferation or fertility.

XX Claim 15; Page 34-35; 69pp; French.

XX The present sequence represents a truncated human kinl7 protein with
 CC amino acids 129-228 deleted. The mammalian kinl7 protein is useful for
 CC preparing a medicament for controlling cell proliferation or for
 CC controlling fertility. The medicaments can also be used to treat
 CC hyperproliferative diseases. Fragments between amino acids 55 and 235
 CC (preferably between amino acids 129 and 228) of a mammalian kinl7 protein
 CC are useful for regulating the interaction between proteins and curved
 CC DNA. The fragment can be used to block replication of HIV or its
 CC integration into the human genome or to target repair enzymes to curved
 CC DNA sites. Expression vectors for kinl7 can be used for controlling cell
 CC proliferation

XX Sequence 293 AA;

```

Query Match          30.7%; Score 120; DB 2; Length 293;
Best Local Similarity 100.0%; Pred. No. 4.3e-107;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  9 PKAIANRIKSGLOKLRYWCQMCQKQCRDENGFKCHCMSESHQROLILLASENPQQFMDYF 68
    |||
Db  9 PKAIANRIKSGLOKLRYWCQMCQKQCRDENGFKCHCMSESHQROLILLASENPQQFMDYF 68
    |||

Qy  69 SEEFNRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATQWETLDTFTKWLREGGLC 128
    |||
Db  69 SEEFNRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATQWETLDTFTKWLREGGLC 128
    |||

RESULT 6
ADP22452
ID  ADP22452 standard; protein; 390 AA.
AC  ADP22452;
DT  12-AUG-2004 (first entry)
XX  Sea-squirt (Ciona intestinalis) zinc finger protein #13.
DE  sea-squirt; zinc finger protein; gene detection; drug development;
KW  zinc finger protein-associated disease.
XX  Ciona intestinalis.
OS  JP2004057126-A.
PN  26-FEB-2004.
XX  31-JUL-2002; 2002JP-00222484.
XX  31-JUL-2002; 2002JP-00222484.
XX  (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX  WPI; 2004-208711/20.
DR  N-PSDB; ADP22451.
XX  Novel gene encoding zinc finger protein, useful as probe in gene
PT  detecting instruments and in development of drug for treating zinc finger
PT  protein associated diseases.
XX  Claim 1; SEQ ID NO 26; 972bp; Japanese.
XX  The invention comprises the amino acid and coding sequences of sea-squirt
CC  (Ciona intestinalis) zinc finger proteins. The DNA and protein sequences
CC  of the invention are useful in a gene detecting instrument. The DNA and
CC  protein sequences of the invention are useful in the development of drugs
CC  for the treatment of zinc finger protein-associated diseases. The present
CC  amino acid sequence represents a sea-squirt zinc finger protein of the
CC  invention.
XX  Sequence 390 AA;

Query Match          10.7%; Score 42; DB 8; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.8e-31;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  14 NRIKSGLOKLRYWCQMCQKQCRDENGFKCHCMSESHQROL 55
    |||
Db  16 NRIKSGLOKLRYWCQMCQKQCRDENGFKCHCMSESHQROL 57
    |||

RESULT 7
AAG42576
ID  AAG42576 standard; protein; 411 AA.
XX  AAG42576;
AC  AAG42576;
XX  Arabidopsis thaliana protein fragment SEQ ID NO: 53114.
XX  Protein identification; signal transduction pathway; metabolic pathway;
KW  hybridisation assay; genetic mapping; gene expression control; promoter;
KW  termination sequence.
XX  Arabidopsis thaliana.
OS  EP1033405-A2.
PN  06-SEP-2000.
PD  25-FEB-2000; 2000EP-00301439.
XX  25-FEB-1999; 99US-0121825P.
XX  05-MAR-1999; 99US-0123180P.
PR  09-MAR-1999; 99US-0123548P.
PR  23-MAR-1999; 99US-0125788P.
PR  25-MAR-1999; 99US-0126264P.
PR  29-MAR-1999; 99US-0126785P.
PR  01-APR-1999; 99US-0127462P.
PR  06-APR-1999; 99US-0128234P.
PR  08-APR-1999; 99US-0128714P.
PR  16-APR-1999; 99US-0129845P.
PR  19-APR-1999; 99US-0130077P.
PR  21-APR-1999; 99US-0130449P.
PR  23-APR-1999; 99US-0130510P.
PR  28-APR-1999; 99US-0130891P.
PR  30-APR-1999; 99US-0131443P.
PR  30-APR-1999; 99US-0132048P.
PR  04-MAY-1999; 99US-0132407P.
PR  05-MAY-1999; 99US-0132484P.
PR  06-MAY-1999; 99US-0132486P.
PR  07-MAY-1999; 99US-0132487P.
PR  11-MAY-1999; 99US-0132863P.
PR  14-MAY-1999; 99US-0134256P.
PR  14-MAY-1999; 99US-0134218P.
PR  14-MAY-1999; 99US-0134219P.
PR  14-MAY-1999; 99US-0134221P.
PR  18-MAY-1999; 99US-0134370P.
PR  18-MAY-1999; 99US-0134768P.
PR  19-MAY-1999; 99US-0134941P.
PR  20-MAY-1999; 99US-0135124P.
PR  21-MAY-1999; 99US-0135353P.
PR  24-MAY-1999; 99US-0135629P.
PR  25-MAY-1999; 99US-0136021P.
PR  27-MAY-1999; 99US-0136392P.
PR  28-MAY-1999; 99US-0136782P.
PR  01-JUN-1999; 99US-0137222P.
PR  03-JUN-1999; 99US-0137528P.
PR  04-JUN-1999; 99US-0137502P.
PR  07-JUN-1999; 99US-0137724P.
PR  08-JUN-1999; 99US-0138094P.
PR  10-JUN-1999; 99US-0138540P.
PR  10-JUN-1999; 99US-0138847P.
PR  14-JUN-1999; 99US-0139119P.
PR  16-JUN-1999; 99US-0139452P.
PR  16-JUN-1999; 99US-0139453P.
PR  17-JUN-1999; 99US-0139492P.
PR  18-JUN-1999; 99US-0139454P.
PR  18-JUN-1999; 99US-0139455P.
PR  18-JUN-1999; 99US-0139456P.
PR  18-JUN-1999; 99US-0139457P.
PR  18-JUN-1999; 99US-0139458P.
PR  18-JUN-1999; 99US-0139459P.
PR  18-JUN-1999; 99US-0139460P.
PR  18-JUN-1999; 99US-0139461P.
PR  18-JUN-1999; 99US-0139462P.
PR  18-JUN-1999; 99US-0139463P.
PR  18-JUN-1999; 99US-0139750P.
```


PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140891P.
PR 29-JUN-1999; 99US-0141287P.
PR 30-JUN-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.

PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 16-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 9.0%; Score 35; DB 3; Length 411;
Best Local Similarity 100.0%; Pred. No. 1.1e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 KGLQLRWYCMQCKQCRDENGFKCHCMSESQRQ 53
Db 19 KGLQLRWYCMQCKQCRDENGFKCHCMSESQRQ 53

RESULT 8
AAB03064
ID AAB03064 standard; protein; 423 AA.
XX
AC AAB03064;
XX
DT 27-SEP-2000 (first entry)
XX
DE Maize KIN17 orthologue, ZmKINH-2.
XX
KW ZmKINH-2; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
nuclear localisation; nonhomologous recombination;
illegitimate recombination; double stranded DNA binding; curved DNA;
KW

```

KW homologous gene targeting; transgenic plant.
XX
OS
XX
XX Zea mays.
XX WO200024900-A1.
XX PN
XX
XX PD
XX 04-MAY-2000.
XX PF
XX 06-OCT-1999; 99WO-US023280.
XX PR
XX 27-OCT-1998; 98US-0105802P.
XX PA
XX (PION-) PIONEER HI-BRED INT INC.
XX PI
XX Mahajan PB;
XX WIPI; 2000-350741/30.
XX DR
XX N-PSDB; AAA52591.
XX
XX Nucleic acids encoding maize KIN17 orthologue proteins useful for
PT preventing illegitimate recombination in cells.
XX
XX Claim 11; Page 64-66; 84pp; English.
XX
XX This sequence represents the maize KIN17 orthologue ZmKINH-2. The
CC invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic
CC acids encoding them (AAA52589-A52591), and expression vectors, transgenic
CC plants and plant seeds comprising nucleotides encoding maize KIN17
CC orthologues. KIN17 has, until now, been found only in animal (avian,
CC rodent and human) cells, this invention being the first report describing
CC the presence of KIN17 in plants. Murine KIN17 was found to have
CC a significant homology to Escherichia coli RecA protein, and contains a
CC zinc finger motif and a nuclear localisation signal. KIN17 binds double-
CC stranded DNA, preferentially binding to curved DNA, and forms
CC intranuclear foci on overexpression in mammalian cells. It is also
CC induced on exposure to gamma or ultraviolet radiation. These findings
CC indicate that KIN17 plays a role in non-homologous (illegitimate)
CC recombination, which occurs at higher rates among higher eukaryotes,
CC particularly plants. Illegitimate recombination in plants is a major
CC impediment to the generation of transgenic crops such as maize. Maize
CC KIN17 orthologue nucleic acid sequences may be used to generate
CC transgenic plants. The transgenic plants generated can be monocots or
CC dicots and are particularly maize, soybean, sunflower, sorghum, canola,
CC wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
CC sequences may be used to reduce KIN17 levels in embryogenic callus or
CC embryo cells, thereby reducing the amount of non-homologous recombination
CC and enhancing homologous gene targeting
XX
XX Sequence 423 AA;
SQ
Query Match 9.0%; Score 35; DB 3; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.1e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53
DB 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53
RESULT 9
AAB03065
ID AAB03065 standard; protein; 423 AA.
XX
XX AAB03065;
XX
XX 27-SEP-2000 (first entry)
XX
XX Maize KIN17 orthologue, ZmKINH-3.
XX
XX ZmKINH-3; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
KW nuclear localisation; nonhomologous recombination;
KW illegitimate recombination; double stranded DNA binding; curved DNA;
KW homologous gene targeting; transgenic plant.
KW

```

```

XX
OS
XX
XX Zea mays.
XX WO200024900-A1.
XX PN
XX
XX PD
XX 04-MAY-2000.
XX PF
XX 06-OCT-1999; 99WO-US023280.
XX PR
XX 27-OCT-1998; 98US-0105802P.
XX PA
XX (PION-) PIONEER HI-BRED INT INC.
XX PI
XX Mahajan PB;
XX WIPI; 2000-350741/30.
XX DR
XX N-PSDB; AAA52591.
XX
XX Nucleic acids encoding maize KIN17 orthologue proteins useful for
PT preventing illegitimate recombination in cells.
XX
XX Claim 11; Page 64-66; 84pp; English.
XX
XX This sequence represents the maize KIN17 orthologue ZmKINH-3. The
CC invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic
CC acids encoding them (AAA52589-A52591), and expression vectors, transgenic
CC plants and plant seeds comprising nucleotides encoding maize KIN17
CC orthologues. KIN17 has, until now, been found only in animal (avian,
CC rodent and human) cells, this invention being the first report describing
CC the presence of KIN17 in plants. Murine KIN17 was found to have
CC a significant homology to Escherichia coli RecA protein, and contains a
CC zinc finger motif and a nuclear localisation signal. KIN17 binds double-
CC stranded DNA, preferentially binding to curved DNA, and forms
CC intranuclear foci on overexpression in mammalian cells. It is also
CC induced on exposure to gamma or ultraviolet radiation. These findings
CC indicate that KIN17 plays a role in non-homologous (illegitimate)
CC recombination, which occurs at higher rates among higher eukaryotes,
CC particularly plants. Illegitimate recombination in plants is a major
CC impediment to the generation of transgenic crops such as maize. Maize
CC KIN17 orthologue nucleic acid sequences may be used to generate
CC transgenic plants. The transgenic plants generated can be monocots or
CC dicots and are particularly maize, soybean, sunflower, sorghum, canola,
CC wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
CC sequences may be used to reduce KIN17 levels in embryogenic callus or
CC embryo cells, thereby reducing the amount of non-homologous recombination
CC and enhancing homologous gene targeting
XX
XX Sequence 423 AA;
SQ
Query Match 9.0%; Score 35; DB 3; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.1e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53
DB 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53
RESULT 10
ADT58422
ID ADT58422 standard; protein; 423 AA.
XX
XX ADT58422;
XX
XX 13-JAN-2005 (first entry)
XX
XX Plant polypeptide, SEQ ID 8499.
XX
XX Plant; transgenic; cold tolerance; growth rate; drought tolerance;
KW disease resistance; galactomannan production; plant growth regulator;
KW heat tolerance; herbicide tolerance; lignin production;
KW extreme osmotic condition tolerance; pathogens resistance;
KW pest resistance; yield improvement; seed oil yield; seed protein yield.
KW

```

XX OS Viridiplantae.
XX PN US2004216190-A1.
XX PD 28-OCT-2004.
XX PF 18-DEC-2003; 2003US-00739930.
XX PR 28-APR-2003; 2003US-00424599.
XX PR 28-APR-2003; 2003US-00425115.
XX PA (KOVA/) KOVALIC D K.
XX PI Kovalic DK;
XX PN WPI; 2004-757369/74.
XX PD New recombinant DNA constructs useful in the field of biochemistry and
XX PT genetics, and in particular for producing transgenic plants with improved
XX PT biological characteristics.
XX PS Claim 2; SEQ ID NO 8499; 14pp; English.
XX CC The invention relates a recombinant DNA construct comprising a
XX CC polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:
XX CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences
XX CC (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,
XX CC Arabidopsis, wheat and rape but the specification does not indicate which
XX CC sequences is derived from which organism. Also included is a method of
XX CC producing a plant having an improved property, comprising transforming a
XX CC plant with a recombinant DNA construct comprising a promoter region
XX CC functional in a plant cell operably joined to a polynucleotide encoding a
XX CC polypeptide associated with the property, and growing the transformed
XX CC plant. The property is selected from improving plant cold tolerance, for
XX CC manipulating growth rate in plant cells by modification of the cell cycle
XX CC pathway, for improving plant drought tolerance, for providing increased
XX CC resistance to plant disease, for galactomannan production, for production
XX CC of plant growth regulators, for improving plant heat tolerance, for
XX CC improving plant tolerance to herbicides, for increasing the rate of
XX CC homologous recombination in plants, for lignin production, for improving
XX CC plant tolerance to extreme osmotic conditions, for improving plant
XX CC tolerance to pathogens or pests, for yield improvement by modification of
XX CC photosynthesis, for modifying seed oil yield and/or content, for
XX CC modifying seed protein yield and/or content, for yield improvement by
XX CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake
XX CC and for yield improvement by providing improved plant growth and
XX CC development under at least one stress condition. The polynucleotide may
XX CC also encode a plant transcription factor. The methods and compositions of
XX CC the present invention are useful in the field of biochemistry and
XX CC genetics, in particular for producing transgenic plants with improved
XX CC biological characteristics such as increased yield, improved nitrogen
XX CC flow, increasing plant tolerance to cold or heat, improving plant
XX CC tolerance to extreme osmotic and drought conditions, and improving plant
XX CC tolerance to plant pests or pathogens. They can also be used in physical
XX CC arrays of molecules, plant breeding markers, computer-based storage and
XX CC analysis systems. The present sequence is one of the 5544 plant protein
XX CC sequences of the invention. Note: The sequence data for this patent did
XX CC not form part of the printed specification, but was obtained in
XX CC electronic format directly from USPTO at
XX CC seqdata.uspto.gov/sequence.html?docID=20040216190.
XX SQ Sequence 423 AA;

Query Match 9.0%; Score 35; DB 8; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.1e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQHQ 53
Db 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQHQ 53

RESULT 11
AAB03063
ID AAB03063 standard; protein; 424 AA.
XX AC AAB03063;
XX DT 27-SEP-2000 (first entry)
XX DE Maize KIN17 orthologue, ZmKINH-1.
XX KW ZmKINH-1; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
XX KW nuclear localisation; nonhomologous recombination;
XX KW illegitimate recombination; double stranded DNA binding; curved DNA;
XX KW homologous gene targeting; transgenic plant.
XX OS Zea mays.
XX PN WO200024900-A1.
XX PD 04-MAY-2000.
XX PF 06-OCT-1999; 99WO-US023280.
XX PR 27-OCT-1998; 98US-0105802P.
XX PR (PION-) PIONEER HI-BRED INT INC.
XX PI Mahajan PB;
XX DR WPI; 2000-350741/30.
XX DR N-PSDB; AAA52589.
XX PT Nucleic acids encoding maize KIN17 orthologue proteins useful for
XX PT preventing illegitimate recombination in cells.
XX PS Claim 11; Page 64-66; 84pp; English.

This sequence represents the maize KIN17 orthologue ZmKINH-1. The
invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic
acids encoding them (AAA52589-A52591), and expression vectors, transgenic
plants and plant seeds comprising nucleotides encoding maize KIN17
orthologues. KIN17 has, until now, been found only in animal (avian,
rodent and human) cells, this invention being the first report describing
the presence of KIN17 in plants. Murine KIN17 was found to have
a significant homology to Escherichia coli RecA protein, and contains a
zinc finger motif and a nuclear localisation signal. KIN17 binds double-
stranded DNA, preferentially binding to curved DNA, and forms
intranuclear foci on overexpression in mammalian cells. It is also
induced on exposure to gamma or ultraviolet radiation. These findings
indicate that KIN17 plays a role in non-homologous (illegitimate)
recombination, which occurs at higher rates among higher eukaryotes,
particularly plants. Illegitimate recombination in plants is a major
impediment to the generation of transgenic crops such as maize. Maize
KIN17 orthologue nucleic acid sequences may be used to generate
transgenic plants. The transgenic plants generated can be monocots or
dicots and are particularly maize, soybean, sunflower, sorghum, canola,
wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
sequences may be used to reduce KIN17 levels in embryogenic callus or
embryo cells, thereby reducing the amount of non-homologous recombination
and enhancing homologous gene targeting

XX SQ Sequence 424 AA;
Query Match 9.0%; Score 35; DB 3; Length 424;
Best Local Similarity 100.0%; Pred. No. 1.1e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQHQ 53
Db 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQHQ 53
RESULT 12

ADX96318
ID ADX96318 standard; protein; 437 AA.
XX
AC ADX96318;
XX
DT 21-APR-2005 (first entry)
XX
DE Plant full length insert polypeptide seqid 58982.
XX
KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content.
XX
OS Unidentified.
XX
PN US2004034888-A1.
XX
PD 19-FEB-2004.
XX
PF 28-APR-2003; 2003US-00425114.
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
PA (LIUJ/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOY/) CAO Y.
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX
XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
PS Claim 1; SEQ ID NO 58982; 15pp; English.
XX
CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This is the amino acid sequence of a plant full length insert
CC polypeptide that can be used in the recombinant DNA construct of the
CC invention.
XX
SQ Sequence 437 AA;
Query Match 9.0%; Score 35; DB 8; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.2e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

||||| 32 KGLQKLRYCQMCKQCRDENGFKCHCMSESHQRQ 66
|||||
DB
RESULT 13
ADX78817
ID ADX78817 standard; protein; 437 AA.
XX
AC ADX78817;
XX
DT 21-APR-2005 (first entry)
XX
DE Plant full length insert polypeptide seqid 48183.
XX
KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content.
XX
OS Unidentified.
XX
PN US2004034888-A1.
XX
PD 19-FEB-2004.
XX
PF 28-APR-2003; 2003US-00425114.
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
PA (LIUJ/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOY/) CAO Y.
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX
XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
PS Claim 1; SEQ ID NO 48183; 15pp; English.
XX
CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This is the amino acid sequence of a plant full length insert
CC polypeptide that can be used in the recombinant DNA construct of the
CC invention.
XX
SQ Sequence 437 AA;

Query Match 9.0%; Score 35; DB 8; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.2e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGIQKLRLWYCMQKQCRDENGPKCHCMSESHQRQ 53
DB 32 KGIQKLRLWYCMQKQCRDENGPKCHCMSESHQRQ 66

RESULT 14
AAG74690
ID AAG74690 standard; protein; 84 AA.
XX
AC AAG74690;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:5454.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.

XX
PN W0200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.

XX 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI; 2001-235357/24.
DR N-PSDB; AAH34095.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.

PS Claim 11; Page 7064-7065; 9803pp; English.

XX
CC AAH32943 to AAH37195 and AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922

SQ Sequence 84 AA;

Query Match 7.4%; Score 29; DB 4; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.6e-19;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 FSATIVETGPKGRRVGGIYEDISKLA 391
DB 56 FSATIVETGPKGRRVGGIYEDISKLA 84

RESULT 15
AAG42577
ID AAG42577 standard; protein; 382 AA.
XX
AC AAG42577;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 53115.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 08-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: November 23, 2005, 16:39:45 ; Search time 24.9362 Seconds
(without alignments)
1508.679 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 391
Sequence: 1 MGKSDLSPKAIANRIKSG.....GPLKGRVREGIQYEDISKLA 391

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 80.*
1: Pirl.*
2: Pirl.*
3: Pirl.*
4: Pirl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	367	93.9	391	2 S18666	KIN17 protein - mo
2	35	9.0	411	2 H96596	hypothetical prote
3	10	2.6	404	2 T27106	hypothetical prote
4	9	2.3	304	2 T40316	conserved kin7-lik
5	9	2.3	3144	2 A46068	Huntington disease
6	8	2.0	303	1 S75983	hypothetical prote
7	8	2.0	342	2 A48258	dopamine receptor
8	8	2.0	400	2 G00013	D3 dopamine recept
9	8	2.0	400	2 G01977	d3 dopamine recept
10	8	2.0	446	1 DVRT03	dopamine receptor
11	8	2.0	446	2 I48322	dopamine receptor
12	8	2.0	461	2 F70571	hypothetical glyci
13	8	2.0	514	2 D56498	probable O-GlcNAc
14	8	2.0	802	2 C94733	hypothetical prote
15	8	2.0	808	2 G86185	endopeptidase La h
16	8	2.0	937	1 S42366	endopeptidase La h
17	8	2.0	962	1 S73342	endopeptidase La h
18	8	2.0	1901	2 F70806	hypothetical glyci
19	7	1.8	57	2 A12525	transposase asr738
20	7	1.8	70	2 B81029	hypothetical prote
21	7	1.8	89	2 AF2709	hypothetical prote
22	7	1.8	89	2 G97491	hypothetical prote
23	7	1.8	100	2 D82245	methylated-DNA-pro
24	7	1.8	121	2 D75089	hypothetical prote
25	7	1.8	136	2 AE2542	transcription repr
26	7	1.8	138	1 Q0BE82	UL73 glycoprotein
27	7	1.8	144	2 C81093	hypothetical prote
28	7	1.8	150	2 A55209	H transfer determi
29	7	1.8	189	2 A80744	conserved hypothet

hypothetical prote
hypothetical prote
probable transaldo
probable transaldo
probable transaldo
transaldolase-like
phosphoglyceromuta
hypothetical prote
vancomycin resista
hypothetical prote
hypothetical prote
spectrin alpha cha
hypothetical prote
merozoite surface
merozoite surface
hypothetical prote

ALIGNMENTS

RESULT 1

S18666
KIN17 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-Feb-1994 #sequence revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S18666; A43753; S14622
R;Angulo, J.F.; Rouer, E.; Mazin, A.; Mattei, M.G.; Tissier, A.; Horellou, P.; Benarous, R.; Angulo, J.; Rouer, E.; Biochimie 73, 251-256, 1991
A;Title: Identification and expression of the cDNA of KIN17, a zinc-finger gene located at Nucleic Acids Res. 19, 5117-5123, 1991
A;Reference number: S18666; MUID:92020193; PMID:1923796
A;Accession: S18666
A;Molecule type: mRNA
A;Residues: 1-391 <ANG>
A;Cross-references: UNIPROT:Q8K339; UNIPARC:UPI000017993F; EMBL:X58472
R;Angulo, J.; Rouer, E.; Benarous, R.; Devoret, R.
Biochimie 73, 251-256, 1991
A;Title: Identification of a mouse cDNA fragment whose expressed polypeptide reacts with Nucleic Acids Res. 19, 5117-5123, 1991
A;Reference number: A43753; MUID:91355299; PMID:1715759
A;Accession: A43753
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 74-273 <ANW>
A;Cross-references: UNIPARC:UPI0000179940
C;Superfamily: KIN17 protein

Query Match	93.9%	Score	367	DB	2	Length	391
Best Local Similarity	100.0%	Pred. No.	0				
Matches	367	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	25	RWYCOMKOCRDENGFKCHCMSESHORQLLLASENPQQFMDYFSEFRNDFLELLRRRF	84				
Db	25	RWYCOMKOCRDENGFKCHCMSESHORQLLLASENPQQFMDYFSEFRNDFLELLRRRF	84				
Qy	85	GTRKRVHNNIVYNYVISHREIHNHATOWETLTDTFKWLGREGLCCKVDETPKGMYIQYIDR	144				
Db	85	GTRKRVHNNIVYNYVISHREIHNHATOWETLTDTFKWLGREGLCCKVDETPKGMYIQYIDR	144				
Qy	145	DPETIRROLEKXKKKODLDEEKTAKFIEQVRRGLEGKEQETPVPTELSRNEEBEKKVT	204				
Db	145	DPETIRROLEKXKKKODLDEEKTAKFIEQVRRGLEGKEQETPVPTELSRNEEBEKKVT	204				
Qy	205	FNLNKAGGSAGATTSSSSSLGPSALKLGSASGKKESSOSSAQPAKKKKSALDEIME	264				
Db	205	FNLNKAGGSAGATTSSSSSLGPSALKLGSASGKKESSOSSAQPAKKKKSALDEIME	264				
Qy	265	LEEBKKRTATDQWLPQGIIVVKIITKLGKHYHKKGVKVEIDRYTAVVVKMTDGSORLK	324				
Db	265	LEEBKKRTATDQWLPQGIIVVKIITKLGKHYHKKGVKVEIDRYTAVVVKMTDGSORLK	324				
Qy	325	LDQTHLETVIPAPGKRVLVNGGYRGNEGTLESINEKAFSATIVTETGPKLGRVEGIIQY	384				
Db	325	LDQTHLETVIPAPGKRVLVNGGYRGNEGTLESINEKAFSATIVTETGPKLGRVEGIIQY	384				

QY 385 EDISKLA 391
Db 385 EDISKLA 391

RESULT 2
H96596
hypothetical protein TSA14.13 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2004 #text_change 09-Jul-2004
C:Accession: H96596
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizär, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96596
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-411 <STO>
A:Cross-references: UNIPROT:Q92VU5; UNIPARC:UPI00000A3CC8; GB:AE005173; NID:g4204268; PI
C:Genetics:
A:Gene: TSA14.13
A:Map position: 1
C:Superfamily: KIN17 protein

Query Match 9.0%; Score 35; DB 2; Length 411;
Best Local Similarity 100.0%; Pred. No. 1.8e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLWWCQMKQKQCRDENGFKCHCMSESHQRQ 53
Db 19 KGLQKLWWCQMKQKQCRDENGFKCHCMSESHQRQ 53

RESULT 3
T27106
hypothetical protein Y52B11A.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T27106
R:Lennard, N.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20310
A:Accession: T27106
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-404 <WIL>
A:Cross-references: UNIPROT:Q9XWF2; UNIPARC:UPI0000078B04; EMBL:AL032654; PIDN:CAA21720.
A:Experimental source: clone Y52B11A
C:Genetics:
A:Gene: CESP:Y52B11A.9
A:Map position: 1
A:Introns: 26/2; 102/3; 301/3; 367/3
C:Superfamily: KIN17 protein

Query Match 2.6%; Score 10; DB 2; Length 404;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 CQMCQKQCRD 37
Db 28 CQMCQKQCRD 37

RESULT 4

T40316
conserved kin7-like hypothetical protein SPBC365.09c - fission yeast (Schizosaccharomycet
C:Species: Schizosaccharomycetes pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T40316
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lelaure, V.; Galibert, F.
submitted to the EMBL Data Library, June 1999
A:Reference number: Z21920
A:Accession: T40316
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-304 <WOO>
A:Cross-references: UNIPROT:Q9YX9; UNIPARC:UPI000006B457; EMBL:AL078627; PIDN:CAB44761.1
A:Experimental source: strain 972h; cosmid c365
C:Genetics:
A:Gene: SPDB:SPBC365.09c
A:Map position: 2

Query Match 2.3%; Score 9; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 RDENGFKCH 44
Db 36 RDENGFKCH 44

RESULT 5
A46068
Huntington disease-associated protein - human
C:Species: Homo sapiens (man)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C:Accession: A46068; I54337
R:MacDonald, M.B.; Ambrose, C.M.; Duyao, M.P.; Myers, R.H.; Lin, C.; Srinidhi, L.; Barnes
J.P.; Bates, G.P.; Baxendale, S.; Hummerich, H.; Kirby, S.; North, M.; Youngman, S.; Mott
O'Donovan, M.C.; Riba-Ramirez, L.; Shah, M.; Stanton, V.P.; Strobel, S.A.; Draths, K.M.
Cell 72, 971-983, 1993
A:Authors: Wales, J.L.; Dervan, P.; Housman, D.E.; Altherr, M.; Shiang, R.; Thompson, L.;
d, K.; Collins, F.S.; Snell, R.; Holloway, T.; Gillespie, K.; Datson, N.; Shaw, D.; Harpe
A:Title: A novel gene containing a trinucleotide repeat that is expanded and unstable on
A:Reference number: A46068; MUID:93208892; PMID:8458085
A:Accession: A46068
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3144 <MAC>
A:Cross-references: UNIPROT:P42858; UNIPARC:UPI0000146F05; GB:U12392
R:Lin, B.; Rommens, J.M.; Graham, R.K.; Kalchman, M.; MacDonald, H.; Nasir, J.; Delaney,
Hum. Mol. Genet. 2, 1541-1545, 1993
A:Title: Differential 3' polyadenylation of the Huntington disease gene results in two m
A:Reference number: I54337; MUID:94093536; PMID:7903579
A:Accession: I54337
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 2563-3144 <RES>
A:Cross-references: UNIPARC:UPI000016AABF; GB:L20431; NID:g398028; PIDN:AAA52702.1; PID:g
C:Genetics:
A:Gene: GDB:HD
A:Cross-references: GDB:119307; OMIM:143100
A:Map position: 4p16.3-4p16.3

Query Match 2.3%; Score 9; DB 2; Length 3144;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
Db 1222 TTSKSSSLG 1230

RESULT 6
S75983
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.

A:Variety: PCC 6803
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
R:Accession: S75983
C:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpō, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
S.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S75983
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <RAN>
A:Cross-references: UNIPROT:Q55483; UNIPARC:UPI000000C0EFA; EMBL:D64006; GB:AB001339; NID submitted to the EMBL Data Library, June 1996
C:Superfamily: hypothetical protein ybbp

Query Match 2.0%; Score 8; DB 1; Length 303;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 76 FLELRRR 83
|||
Db 100 FLELRRR 107

RESULT 7
A48258
dopamine receptor variant D3nf (schizophrenia) - human
C:Species: Homo sapiens (man)
C:Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 09-Jul-2004
R:Schmauss, C.; Haroutunian, V.; Davis, K.L.; Davidson, M.
Proc. Natl. Acad. Sci. U.S.A. 90, 8942-8946, 1993
A:Title: Selective loss of dopamine D3-type receptor mRNA expression in parietal and motor cortex in schizophrenia
A:Reference number: A48258; MUID:94022291; PMID:8415635
A:Accession: A48258
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-342 <RES>
A:Cross-references: UNIPROT:P35462; UNIPARC:UPI0000002A7A9; GB:L20469; NID:g306688; PIDN:
C:Superfamily: vertebrate rhodopsin
C:Keywords: neurotransmitter receptor

Query Match 2.0%; Score 8; DB 2; Length 342;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 307 IDRYTAVV 314
|||
Db 126 IDRYTAVV 133

RESULT 8
G00013
D3 dopamine receptor - green monkey
C:Species: Cercopithecus aethiops (green monkey, grivet)
C:Date: 13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C:Accession: G00013
R:Ross, P.C.
submitted to the EMBL Data Library, February 1995
A:Reference number: G00049
A:Accession: G00013
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-400 <ROS>
A:Cross-references: UNIPROT:P52703; UNIPARC:UPI00000128DB0; EMBL:U21307; NID:g984965; PID
C:Superfamily: vertebrate rhodopsin
C:Keywords: neurotransmitter receptor

Query Match 2.0%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 307 IDRYTAVV 314

Db 126 IDRYTAVV 133

RESULT 9

G01977

d3 dopamine receptor - human

C:Species: Homo sapiens (man)

C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004

C:Accession: G01977

R:Fishburn, C.S.; Park, B.

submitted to the EMBL Data Library, July 1995

A:Reference number: G08971

A:Accession: G01977

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-400 <FIS>

A:Cross-references: UNIPROT:P35462; UNIPARC:UPI0000002CDFB; EMBL:U32499; NID:g927341; PIDN:

C:Superfamily: vertebrate rhodopsin

Query Match 2.0%; Score 8; DB 2; Length 400;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 307 IDRYTAVV 314

Db 126 IDRYTAVV 133

RESULT 10

DYRTD3

dopamine receptor D3 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004

C:Accession: S11565; S41849; I52280

R:Sokoloff, P.; Giros, B.; Martres, M.P.; Bouthenet, M.L.; Schwartz, J.C.

Nature 347, 146-151, 1990

A:Title: Molecular cloning and characterization of a novel dopamine receptor (D3) as a

A:Reference number: S11565; MUID:90370111; PMID:1975644

A:Accession: S11565

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-446 <SOK>

A:Cross-references: UNIPROT:P19020; UNIPARC:UPI0000002CB23; EMBL:X53944; NID:g56060; PIDN:

R:Giros, B.; Martres, M.P.; Pilon, C.; Sokoloff, P.; Schwartz, J.C.

Biochem. Biophys. Res. Commun. 176, 1584-1592, 1991

A:Title: Shorter variants of the D(3) dopamine receptor produced through various pattern

A:Reference number: S41849; MUID:91248260; PMID:2039532

A:Accession: S41849

A:Molecule type: DNA

A:Residues: 1-138, 'E', 140-446 <GIR>

A:Cross-references: UNIPARC:UPI0000002CB22

R:Pagliusi, S.; Chollet-Daemmerius, A.; Losberger, C.; Mills, A.; Kawashima, E.

Biochem. Biophys. Res. Commun. 194, 465-471, 1993

A:Title: Characterization of a novel exon within the D3 receptor gene giving rise to an

A:Reference number: I52280; MUID:93326159; PMID:8333859

A:Accession: I52280

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 71-106 <RES>

A:Cross-references: UNIPARC:UPI00000170C6A; GB:S63847; NID:g399705; PIDN:AAB27545.1; PID:

A:Experimental source: strain Wistar, brain

C:Genetics:

A:Introns: 90/3

C:Superfamily: vertebrate rhodopsin

C:Keywords: G protein-coupled receptor; glycoprotein; lipoprotein; neurotransmitter rece

F:33-55/Domain: transmembrane #status predicted <TM1>

F:67-92/Domain: transmembrane #status predicted <TM2>

F:105-126/Domain: transmembrane #status predicted <TM3>

F:150-172/Domain: transmembrane #status predicted <TM4>

F:186-209/Domain: transmembrane #status predicted <TM5>

F;210-375/Domain: intracellular #status predicted <CYT>
F;376-399/Domain: transmembrane #status predicted <TM6>
F;413-434/Domain: transmembrane #status predicted <TM7>
F;12,19/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;103-181/Disulfide bonds: #status predicted
F;446/Binding site: palmitate (Cys) (covalent) #status predicted

Query Match 2.0%; Score 8; DB 1; Length 446;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133
|||||

RESULT 11
I48322
dopamine receptor D3 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C:Accession: I48322; S25161
R:Fishburn, C.S.; Belleli, D.; David, C.; Carmon, S.; Fuchs, S.
J. Biol. Chem. 268, 5872-5878, 1993
A:Title: A novel short isoform of the D3 dopamine receptor generated by alternative splicing
A:Reference number: I48322; MUID:93194894; PMID:8449953
A:Accession: I48322
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-446 <RES>
A:Cross-references: UNIPROT:P30728; UNIPARC:UPI0000021PB7; EMBL:X67274; NID:950651; PTDN
C:Superfamily: vertebrate rhodopsin
C:Keywords: neurotransmitter receptor

Query Match 2.0%; Score 8; DB 2; Length 446;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133
|||||

RESULT 12
F70571
hypothetical glycine-rich protein RV2615c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: F70571
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70571
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-461 <COL>
A:Cross-references: UNIPROT:O06199; UNIPARC:UPI0000165306; GB:Z95387; GB:AL123456; NID:9
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV2615c
C:Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 2.0%; Score 8; DB 2; Length 461;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GAGGSAGA 217
Db 379 GAGGSAGA 386
|||||

RESULT 13
D56849
dopamine receptor-like protein D222 - Japanese pufferfish
C:Species: Fugu rubripes (Japanese pufferfish)
C:Date: 27-Oct-1995 #sequence_revision 27-Oct-1995 #text_change 09-Jul-2004
C:Accession: D56849
R:Macrae, A.D.; Brenner, S.
Genomics 25, 436-446, 1995
A:Title: Analysis of the dopamine receptor family in the compact genome of the puffer fish
A:Reference number: A56849; MUID:95309511; PMID:7789977
A:Accession: D56849
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-514 <MAC>
A:Cross-references: UNIPROT:O9PSA6; UNIPARC:UPI00000FBA2D
C:Superfamily: vertebrate rhodopsin
C:Keywords: neurotransmitter receptor

Query Match 2.0%; Score 8; DB 2; Length 514;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 131 IDRYTAVV 138
|||||

RESULT 14
C84733
probable O-GlcNAc transferase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: C84733
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; A.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C84733
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <STO>
A:Cross-references: UNIPROT:Q8S8L9; UNIPARC:UPI000000C2F7; GB:AE002093; NID:96598434; PII
C:Genetics:
A:Gene: At2g32450
A:Map position: 2

Query Match 2.0%; Score 8; DB 2; Length 802;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 ALKILGSA 236
Db 341 ALKILGSA 348
|||||

RESULT 15
G86185
hypothetical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: G86185
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luross, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, J.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86185
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-808 <STO>
A:Cross-references: UNIPROT:O23052; UNIPARC:UPI00000481C8; GB:AE005172; NID:g2388582; PI
C:Genetics:
A:Map position: 1

Query Match 2.0%; Score 8; DB 2; Length 808;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 ALKLLGSA 236
| | | | | | | |
DB 346 ALKLLGSA 353

Search completed: November 23, 2005, 16:51:46
Job time : 26.9362 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:34:39 ; Search time 154.106 Seconds
(without alignments)
1790.078 Million cell updates/sec

Title: US-09-555-529-25

Perfect score: 391

Sequence: 1 MCKSDFLSPKAIANRIKSG.....GPLKGRVREGIQVEDISKLA 391

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2166443 seqs, 705528306 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	391	100.0	391	Q8K339_MOUSE	Q8K339 mus musculus
2	195	49.9	195	Q9CV58_MOUSE	Q9CV58 mus musculus
3	179	45.8	393	Q60870_HUMAN	Q60870 homo sapien
4	61	15.6	389	Q6G197_XENTR	Q6G197 xenopus tro
5	50	12.8	387	Q68F56_XENLA	Q68F56 xenopus lae
6	48	12.3	427	Q4S5G2_TETNG	Q4S5G2 tetraodon n
7	43	11.0	383	Q5RH25_BRARE	Q5RH25 brachydanio
8	41	10.5	269	Q4TDV9_TETNG	Q4TDV9 tetraodon n
9	35	9.0	402	Q9AY87_ORYSA	Q9AY87 oryza sativ
10	35	9.0	411	Q9ZVU5_ARATH	Q9ZVU5 arabidopsis
11	35	9.0	430	Q75LU5_ORYSA	Q75LU5 oryza sativ
12	20	5.1	188	Q4TDW0_TETNG	Q4TDW0 tetraodon n
13	15	3.8	178	Q6X190_DROVA	Q6X190 drosophila
14	15	3.8	244	Q76926_DROME	Q76926 drosophila
15	15	3.8	390	Q8SXH2_DROME	Q8SXH2 drosophila
16	15	3.8	390	Q9VPH4_DROME	Q9VPH4 drosophila
17	14	3.6	263	Q5CFM3_CRYHO	Q5CFM3 cryptospori
18	14	3.6	265	Q5CYD1_CRYPV	Q5CYD1 cryptospori
19	13	3.3	387	Q7PGAL_ANOGA	Q7PGAL anopheles g
20	12	3.1	121	Q4XND8_PLACH	Q4XND8 plasmodium
21	12	3.1	239	Q6CXJ7_KLUFA	Q6CXJ7 kluyveromyc
22	12	3.1	378	Q4N9K7_THEPA	Q4N9K7 theileria p
23	12	3.1	412	Q4UG63_THEAN	Q4UG63 theileria a
24	12	3.1	441	Q4XDV6_PLACH	Q4XDV6 plasmodium
25	12	3.1	442	Q8IKG2_PLAF7	Q8IKG2 plasmodium
26	12	3.1	445	Q5SD16_DICDI	Q5SD16 dictyosteli
27	12	3.1	445	Q4Z7D7_PLABE	Q4Z7D7 plasmodium
28	11	2.8	457	Q7RD49_PLAYO	Q7RD49 plasmodium
29	11	2.8	238	Q51F62_ENTHI	Q51F62 entamoeba h
30	10	2.6	397	Q60L39_CAEBR	Q60L39 caenorhabdi
31	10	2.6	404	Q9XWF2_CABEL	Q9XWF2 caenorhabdi

Q9Y7X9 echizoaecch
Q55M94 cryptococcu
Q5KH7 cryptococcu
Q4WGO aspergillus
Q5A80 aspergillus
Q59F4 homo sapien
P42B58 homo sapien
Q6PPZ4 plasmodium
O18876 macaca fasc
Q6PPZ4 plasmodium
Q4NX48 anaeromyxob
Q9F44 oryza sativ
Q76218 oncorhynchu
Q8XVB3 ralstonia s

32 9 2.3 304 2 Q9Y7X9 SCHPO
33 9 2.3 324 2 Q55M94 CRYNE
34 9 2.3 324 2 Q5KH7 CRYNE
35 9 2.3 334 2 Q4WGO ASPFU
36 9 2.3 335 2 Q5A80 EMENI
37 9 2.3 1147 2 Q59F4 HUMAN
38 9 2.3 3144 1 HD HUMAN
39 8 2.0 54 2 Q6PPZ4 PLAF7
40 8 2.0 67 2 O18876 MACFA
41 8 2.0 131 2 Q6PPZ4 PLAF7
42 8 2.0 134 2 Q4NX48 DELT
43 8 2.0 143 2 Q9FRF4 ORYSA
44 8 2.0 158 2 Q76218 ONCNE
45 8 2.0 174 2 Q8XVB3 RALSO

ALIGNMENTS

RESULT 1

Q8K339_MOUSE

ID Q8K339_MOUSE PRELIMINARY; PRT; 391 AA.

AC Q8K339; DT 01-OCT-2002 (TREMBlrel. 22, Created)

DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)

DE Antigenic determinant of rec-A protein.

GN Name=Kin;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Brownstein M.J., Usdin T.B., Toshyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;

RA Strausberg R.; Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

RL [3]

RP NUCLEOTIDE SEQUENCE.

RC MEDLINE=32020193; PubMed=1923796;

RX Angulo J., Rouet E., Mazin A., Mattei M., Tissier A., Horellou P.,

RA Benarous R., Devoret R.;

RT "Identification and expression of the cDNA of KIN17, a zinc-finger

gene located on mouse chromosome 2, encoding a new DNA binding

protein.";

RL Nucleic Acids Res. 19:5117-5123(1991).

DR EMBL; BC028860; AA028860.1; -; mRNA.

DR PIR; S18666; S18666.

```
DR Ensemble; ENSMUSG00000037262; Mus musculus.
DR MGI:96676; Kin.
DR GO:0005634; C:nucleus; IEA.
DR GO:0003676; F:nucleic acid binding; IEA.
DR GO:0008270; F:zinc ion binding; IEA.
DR InterPro: IPR005824; KOW.
DR InterPro: IPR006646; KOW sub.
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00467; KOW; 1.
DR SMART: SM00739; KOW; 1.
DR SMART: SM00355; Znf_C2H2; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nucleic protein.
SQ SEQUENCE 391 AA; 44722 MW; 955BC6A4FF4D3B6E CRC64;

Query Match      100.0%; Score 391; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKSDFLSPKATANRIKSGLOKLRLWYCMQCKQCRDENGFKCHCMSESHQQLLIASEN 60
Db 1 MGKSDFLSPKATANRIKSGLOKLRLWYCMQCKQCRDENGFKCHCMSESHQQLLIASEN 60

QY 61 PQQFMDYFSEERPNDFLELLRRFRGTRVHNNIVVNEYISHREHIMNATOWETLTDFTK 120
Db 61 PQQFMDYFSEERPNDFLELLRRFRGTRVHNNIVVNEYISHREHIMNATOWETLTDFTK 120

QY 121 WLGREGCLKVDTPKGYIYQIDROPETIRRLQLELEKQKQDDEEKTAKTIEEQVRRG 180
Db 121 WLGREGCLKVDTPKGYIYQIDROPETIRRLQLELEKQKQDDEEKTAKTIEEQVRRG 180

QY 181 LEGKQETVPVTELSRENEEEKVTNFKNGAGSGAGATTSKSSSLGSPSALKLLGSAASGK 240
Db 181 LEGKQETVPVTELSRENEEEKVTNFKNGAGSGAGATTSKSSSLGSPSALKLLGSAASGK 240

QY 241 RKESSOSSAQAQAKKKSALDEIMELSEEKRTARTDQWLOPGIVVKIITPKLGEKYHKKK 300
Db 241 RKESSOSSAQAQAKKKSALDEIMELSEEKRTARTDQWLOPGIVVKIITPKLGEKYHKKK 300

QY 301 GVVKEVIDRYTAVVXMTDSGDRLLKLDQTHLETVIPAPGRKRLVNLNGYRGNEGTLLESLNE 360
Db 301 GVVKEVIDRYTAVVXMTDSGDRLLKLDQTHLETVIPAPGRKRLVNLNGYRGNEGTLLESLNE 360

QY 361 KAFSATIVETGPKLGRVREGIQYEDISKLA 391
Db 361 KAFSATIVETGPKLGRVREGIQYEDISKLA 391

RESULT 2
Q9CV58_MOUSE
ID Q9CV58_MOUSE PRELIMINARY; PRT; 195 AA.
AC Q9CV58;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Mus musculus adult male tongue cDNA, RIKEN full-length enriched
DE library, clone:2310020017 product:antigenic determinant of rec-A
DE protein, full insert sequence. (Fragment).
GN Name=Kin;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI TaxID=10090;
RN [1]_
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.

STRAIN=C57BL/6J; TISSUE=Tongue;
MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
Kuehl P., Lewis S., Mateu Y., Nikaide I., Pesole G., Quackenbush J.,
Sakai L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
Brownstein S.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Guscinich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=20530913; PubMed=11078861; DOI=10.1101/gr.152600;
Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=20530913; PubMed=11078861; DOI=10.1101/gr.152600;
Aizawa K., Konno H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
Arakawa T., Bono H., Carninci P., Hiramoto K., Hiraoka T., Hori F.,
Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK009429; BAB26281.2; -, mRNA.
DR MGI; MGI:96676; Kin.
DR GO:0005634; C:nucleus; IEA.
DR GO:0003676; F:nucleic acid binding; IEA.
DR GO:0008270; F:zinc ion binding; IEA.
```

DR InterPro: IPR007087; Znf_C2H2.
 DR SMART: SM00355; Znf_C2H2_1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 FT NON_TER 195
 SQ SEQUENCE 195 AA; 23518 MW; 7E112E9CA820BFA6 CRC64;

Query Match 49.9%; Score 195; DB 2; Length 195;
 Best Local Similarity 100.0%; Pred. No. 1.2e-188;
 Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKSDFLSPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHQROLILASEN 60
 DB 1 MGKSDFLSPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHQROLILASEN 60
 QY 61 PQQPMDFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDTFK 120
 DB 61 PQQPMDFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDTFK 120
 QY 121 WLREGGLCKVDTPKGYIOYIDRDPETIRROLELEKKKKQDLDDEKTAFFIEEQVRRG 180
 DB 121 WLREGGLCKVDTPKGYIOYIDRDPETIRROLELEKKKKQDLDDEKTAFFIEEQVRRG 180
 QY 181 LEGKEQETPVFTLS 195
 DB 181 LEGKEQETPVFTLS 195

RESULT 3

Q60870 HUMAN
 ID O60870 HUMAN PRELIMINARY; PRT; 393 AA.
 AC O60870;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Kin17 protein (HsKin17 protein) (KIN, antigenic determinant of recA protein homolog) (Moues).
 DE Name: KIN; Synonyms: Kin17; ORFNames=RP11-264C14.1-001;
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Choroontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Testis;
 RX MEDLINE=20419742; PubMed=10964102; DOI=10.1093/carcin/21.9.1701;
 RA Kannouche P., Mauffrey P., Pinon-Lataillade G., Mattei M.G.,
 RA Sarasin A., Daya-Grosjean L., Angulo J.F.;
 RA "Molecular cloning and characterization of the human KIN17 cDNA encoding a component of the UV response that is conserved among metazoans.";
 RL Carcinogenesis 21:1701-1710(2000).
 RN [2]

RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haie H.F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahney J., Helton E., Kettelman M., Madan A.C., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Trinchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Groomwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

*Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Pancreas;
 RA Strausberg R.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RA Lovell J.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ005273; CA06462.1; -; mRNA.
 DR EMBL; BC017309; AAH17309.1; -; mRNA.
 DR EMBL; AL158044; CA112959.1; -; Genomic DNA.
 DR Ensembl; ENSG00000151657; Homo sapiens.
 DR HGNC; HGNC:6327; KIN.
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0003677; F:DNA binding; TAS.
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00467; KOW; 1.
 DR SMART; SM00355; Znf_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Nuclear protein.
 SQ SEQUENCE 393 AA; 45374 MW; 515A89B4C8A4C007 CRC64;

Query Match 45.8%; Score 179; DB 2; Length 393;
 Best Local Similarity 100.0%; Pred. No. 3.7e-172;
 Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 PKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHQROLILASENPQPMDF 68
 DB 9 PKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHQROLILASENPQPMDF 68
 QY 69 SESEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDTFKLGRGLC 128
 DB 69 SESEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDTFKLGRGLC 128
 QY 129 KVDETCKGYIOYIDRDPETIRROLELEKKKKQDLDDEKTAFFIEEQVRRGLEKQE 187
 DB 129 KVDETCKGYIOYIDRDPETIRROLELEKKKKQDLDDEKTAFFIEEQVRRGLEKQE 187

RESULT 4
 Q6GL97_XENTR PRELIMINARY; PRT; 389 AA.
 ID Q6GL97_XENTR PRELIMINARY;
 AC Q6GL97;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE MGC69184 protein.
 GN Name=MGC69184;
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus; Silurana.
 OX NCBI_TaxID=8364;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haie H.F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

Query Match 12.3%; Score 48; DB 2; Length 427;
 Best Local Similarity 100.0%; Pred. No. 2.8e-39;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 NRIKSGQLRWYQCMQKQCRDENGFKCHMSHQRQLLASNP 61
 |||||
 DB 14 NRIKSGQLRWYQCMQKQCRDENGFKCHMSHQRQLLASNP 61
 |||||

RESULT 7
 Q5RH25 BRARE PRELIMINARY; PRT; 383 AA.
 AC Q5RH25
 DT 01-FEB-2005 (TrEMBLrel. 29, Created)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 DE Novel protein similar to vertebrate KIN, antigenic determinant of reca
 DE protein homolog (Mouse) (KIN).
 GN Name:OTTARP0000008158; ORFNames=CH211-235A22.3-001;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Hunter G.;
 RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BX640594; CA11855.1; -; Genomic_DNA.
 DR InterPro; IPR005824; K0M.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00467; K0M; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 SQ SEQUENCE 383 AA; 44337 MW; 1F0AA34FB62176D1 CRC64;

Query Match 11.0%; Score 43; DB 2; Length 383;
 Best Local Similarity 100.0%; Pred. No. 3e-34;
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLWYQCMQKQCRDENGFKCHMSHQRQLLASNP 61
 |||||
 DB 19 KGLQKLWYQCMQKQCRDENGFKCHMSHQRQLLASNP 61
 |||||

RESULT 8
 Q4TDV9 TETNG PRELIMINARY; PRT; 269 AA.
 AC Q4TDV9
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Chromosome undetermined SCAF6005, whole genome shotgun sequence.
 DE (Fragment).
 GN ORFNames=GSTENG0002601001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallou O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Bismont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
 RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,

RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Winkler P., Lander E.S., Weissenbach J., Roest Crolius H.;
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL; CAE01006005; CAF88923.1; -; Genomic_DNA.
 FT NON_TER 1 1
 FT TER 269 269
 SQ SEQUENCE 269 AA; 31736 MW; E6883839A2CD17AF CRC64;

Query Match 10.5%; Score 41; DB 2; Length 269;
 Best Local Similarity 100.0%; Pred. No. 2.3e-32;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 WETLTDTFTKWLGRGLCKVDTEPKGWYQYIDRDPETIRQ 152
 |||||
 DB 58 WETLTDTFTKWLGRGLCKVDTEPKGWYQYIDRDPETIRQ 98
 |||||

RESULT 9
 Q9AY87 ORYSA PRELIMINARY; PRT; 402 AA.
 AC Q9AY87
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein OSJNBa0004B24.17.
 GN Names:OSJNBa0004B24.17;
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoidae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,
 RA Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell R.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC084319; AAC59654.1; -; Genomic_DNA.
 DR Gramene; Q9AY87; -.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0046872; F:metal ion binding; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR007087; Znf_C2H2.
 DR SMART; SM00355; Znf_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 402 AA; 45962 MW; A1F4E38ECC60D13E CRC64;

Query Match 9.0%; Score 35; DB 2; Length 402;
 Best Local Similarity 100.0%; Pred. No. 4e-26;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLWYQCMQKQCRDENGFKCHMSHQRQ 53
 |||||
 DB 19 KGLQKLWYQCMQKQCRDENGFKCHMSHQRQ 53
 |||||

RESULT 10

O9ZVU5 ARATH
 ID Q9ZVU5_ARATH PRELIMINARY; PRT; 411 AA.
 AC Q9ZVU5;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 DE T5A14.13 protein (Hypothetical protein At1g55460).
 GN Name=T5A14.13; Synonyms=At1g55460;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
 RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
 RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
 RA Lurse S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
 RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,
 RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.B., Toriumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC005223; AAD10649.1; -; Genomic_DNA.
 DR EMBL; AY051011; AAK93688.1; -; mRNA.
 DR EMBL; AF360132; AAK25842.1; -; mRNA.
 DR PIR; H96596; H96596.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00467; KOW; 1.
 DR SMART; SM00739; KOW; 1.
 DR SMART; SM00355; Znf_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 411 AA; 47288 MW; 9DA6F8648002065D CRC64;

 Q7SLU5 ORYZA
 ID Q7SLU5_ORYZA PRELIMINARY; PRT; 430 AA.
 AC Q7SLU5;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein OSJNB0015121.3.
 GN Names=OSJNB0015121.3;
 RESULT 11
 Q7SLU5 ORYZA
 ID Q7SLU5_ORYZA PRELIMINARY; PRT; 430 AA.
 AC Q7SLU5;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein OSJNB0015121.3.
 GN Names=OSJNB0015121.3;
 RESULT 12
 Q4TDMO_TETNG
 ID Q4TDMO_TETNG PRELIMINARY; PRT; 88 AA.
 AC Q4TDMO;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Chromosome undetermined SCAF6004, whole genome shotgun sequence.
 DE (Fragment).
 OS ORFNames=GSTENG00002600001;
 GN Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Teleostei; Neoteleostei;
 OC Actinopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicod S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salancoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Bionnot C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Winkler P., Lander E.S., Weissbach J., Roest Crolius H.;
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
 RA Overton II L.B., Tsitrin T., Kim M.M., Bera J.O., Jin S.S.,
 RA Fardosh D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,
 RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feidblyum T.V.,
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
 RA White O., Salzberg S.L., Fraser C.M.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell R.;
 RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC091302; AAR00634.1; -; Genomic_DNA.
 DR Gramene; Q75LU5;
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00467; KOW; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 430 AA; 49067 MW; 6151D748ED2F4690 CRC64;

 Query Match 9.0%; Score 35; DB 2; Length 430;
 Best Local Similarity 100.0%; Pred. No. 4.3e-26;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQRQ 53
 DB 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQRQ 53

 RESULT 12
 Q4TDMO_TETNG
 ID Q4TDMO_TETNG PRELIMINARY; PRT; 88 AA.
 AC Q4TDMO;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Chromosome undetermined SCAF6004, whole genome shotgun sequence.
 DE (Fragment).
 OS ORFNames=GSTENG00002600001;
 GN Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Teleostei; Neoteleostei;
 OC Actinopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicod S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salancoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Bionnot C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Winkler P., Lander E.S., Weissbach J., Roest Crolius H.;
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

```
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data!
CC EMBL; CAEE01006004; CAF88922.1; -; Genomic_DNA.
DR
DR
FT NON_TER 1
FT NON_TER 88
SQ SEQUENCE 88 AA; 9709 MW; C5BD169AF081CBC2 CRC64;

Query Match
Best Local Similarity 5.1%; Score 20; DB 2; Length 88;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 323 LKLDQTHLETVIPAPGRVL 342
Db 21 LKLDQTHLETVIPAPGRVL 40

RESULT 13
O6XI90_DROYA
ID Q6XI90_DROYA PRELIMINARY; PRT; 178 AA.
AC Q6XI90;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Similar to Drosophila melanogaster kin17 (Fragment).
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22887302; PubMed=14525923; DOI=10.1101/gr.1311003;
RA Domazet-Looso T., Tautz D.;
RT "An evolutionary analysis of orphan genes in Drosophila.";
RL Genome Res. 13:2213-2219(2003).
DR EMBL; AY231941; AAR09964.1; -; mRNA.
DR InterPro; IPR007087; Znf_C2H2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 178
SQ SEQUENCE 178 AA; 21188 MW; B251B8630CBC76C3 CRC64;

Query Match
Best Local Similarity 3.8%; Score 15; DB 2; Length 178;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 KSKGLQKLRYWCQC 31
Db 17 KSKGLQKLRYWCQC 31

RESULT 14
O76926_DROME
ID O76926_DROME PRELIMINARY; PRT; 244 AA.
AC O76926;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE KIN17 protein (Fragment).
GN Name=kin17; ORFNames=CG5649;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=EST28;
RA Garcia V., de La Roche Saint Andre C.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ006528; CAA07089.1; -; Genomic_DNA.
DR Ensembl; CG5649; Drosophila melanogaster.

FlyBase; FBgn0024887; CG5649.
FlyBase; FBgn0024887; kin17.
GO; GO:0005634; C:nucleus; IEA.
GO; GO:0003676; F:nucleic acid binding; IEA.
GO; GO:0008270; F:zinc ion binding; IEA.
InterPro; IPR007087; Znf_C2H2.
SMART; SM00355; Znf_C2H2.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
NON_TER 244
SQ SEQUENCE 244 AA; 28679 MW; 5AF6C28524234582 CRC64;

Query Match
Best Local Similarity 3.8%; Score 15; DB 2; Length 244;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 KSKGLQKLRYWCQC 31
Db 17 KSKGLQKLRYWCQC 31

RESULT 15
Q8XR2_DROME
ID Q8XR2_DROME PRELIMINARY; PRT; 390 AA.
AC Q8XR2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE R65257P.
GN Name=kin17; ORFNames=CG5649;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champs M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Murgall C.J., Nunco J., Pacleb J., Parages V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celnikier S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY084184; AAL89922.1; -; mRNA.
DR FlyBase; FBgn0024887; kin17.
DR FlyBase; FBgn0024887; CG5649.
GO; GO:0005634; C:nucleus; IEA.
GO; GO:0003676; F:nucleic acid binding; IEA.
GO; GO:0008270; F:zinc ion binding; IEA.
InterPro; IPR007087; Znf_C2H2.
SMART; SM00355; Znf_C2H2.
PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ SEQUENCE 390 AA; 45286 MW; 39590CD3B61077CC CRC64;

Query Match
Best Local Similarity 3.8%; Score 15; DB 2; Length 390;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 KSKGLQKLRYWCQC 31
Db 17 KSKGLQKLRYWCQC 31

Search completed: November 23, 2005, 16:49:37
Job time : 155.106 secs
```

This Page Blank (usps)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:35:49 ; Search time 32.9158 Seconds
(without alignments)
982.087 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 391
Sequence: 1 MCKSDFLSPKAIANRIKSG.....GPKLGRVGIQVEDISKLA 391

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 572060 seqs, 82675679 residues

Word size : 0

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5 COMB.pep.*
2: /cgn2_6/prodata/1/iaa/6 COMB.pep.*
3: /cgn2_6/prodata/1/iaa/H COMB.pep.*
4: /cgn2_6/prodata/1/iaa/PCITUS COMB.pep.*
5: /cgn2_6/prodata/1/iaa/RE COMB.pep.*
6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	2.3	1543	2	US-09-904-987-7
2	9	2.3	3144	1	US-08-246-982A-6
3	9	2.3	3144	1	US-08-453-265-6
4	9	2.3	3144	1	US-08-457-273B-42
5	9	2.3	3144	2	US-08-556-419-21
6	9	2.3	3144	2	US-09-041-886-15
7	9	2.3	3144	2	US-09-538-092-1118
8	8	2.0	136	2	US-09-270-767-32008
9	8	2.0	400	2	US-09-826-509-491
10	8	2.0	444	1	US-07-626-618A-20
11	8	2.0	444	1	US-08-333-977-20
12	8	2.0	446	1	US-07-781-254A-2
13	8	2.0	446	1	US-07-781-254A-3
14	8	2.0	546	2	US-09-252-991A-18637
15	8	2.0	562	2	US-09-949-002-426
16	8	2.0	937	1	US-08-253-155A-31
17	8	2.0	937	2	US-09-538-092-1092
18	8	2.0	937	2	US-09-949-002-374
19	8	2.0	959	2	US-09-538-092-1091
20	7	1.8	58	2	US-07-741-453A-41
21	7	1.8	65	2	US-09-881-572A-1
22	7	1.8	68	2	US-09-270-767-34723
23	7	1.8	68	2	US-09-270-767-49940
24	7	1.8	74	2	US-09-489-039A-7206
25	7	1.8	76	2	US-09-621-976-7070
26	7	1.8	80	2	US-08-930-503A-16
27	7	1.8	90	2	US-09-613-486-24

28 7 1.8 91 2 US-09-134-001C-4276 Sequence 4276, Ap
29 7 1.8 98 2 US-09-489-039A-7316 Sequence 7316, Ap
30 7 1.8 99 2 US-09-866-108A-15755 Sequence 15755, A
31 7 1.8 120 2 US-09-311-021-20 Sequence 20, Appl
32 7 1.8 152 2 US-09-248-796A-15608 Sequence 15608, A
33 7 1.8 197 2 US-09-489-039A-14336 Sequence 14336, A
34 7 1.8 201 2 US-09-744-778-4 Sequence 4, Appl
35 7 1.8 208 2 US-09-252-991A-32430 Sequence 32430, A
36 7 1.8 218 2 US-09-949-016-6586 Sequence 6586, Ap
37 7 1.8 229 2 US-09-904-923C-1 Sequence 1, Appl
38 7 1.8 238 1 US-08-103-170-13 Sequence 13, Appl
39 7 1.8 250 2 US-09-809-665A-83 Sequence 83, Appl
40 7 1.8 253 2 US-09-248-796A-15042 Sequence 15042, A
41 7 1.8 262 2 US-09-489-039A-12195 Sequence 12195, A
42 7 1.8 263 2 US-09-902-540-10610 Sequence 10610, A
43 7 1.8 271 2 US-09-949-016-11720 Sequence 11720, A
44 7 1.8 277 1 US-08-400-413-1 Sequence 1, Appl
45 7 1.8 299 2 US-09-248-796A-15986 Sequence 15986, A

ALIGNMENTS

RESULT 1
US-09-904-987-7
; Sequence 7, Application US/09904987
; Patent No. 6743771
; GENERAL INFORMATION:
; APPLICANT: No. 6743771actyl, Inc.
; TITLE OF INVENTION: Methods and Compositions for Controlling Pathological and Prepatk
; FILE OF INVENTION: Protein Assembly or Aggregation
; FILE REFERENCE: 42108/26146
; CURRENT APPLICATION NUMBER: US/09/904, 987
; CURRENT FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 7
; LENGTH: 1543
; TYPE: PRT
; ORGANISM: homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NCBI ENTREZ / XP_003405
; DATABASE ENTRY DATE: 2001-04-16
; RELEVANT RESIDUES: (1)..(1543)
US-09-904-987-7

Query Match 2.3%; Score 9; DB 2; Length 1543;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 TTSSKSSSLG 226
Db 1219 TTSSKSSSLG 1227

RESULT 2
US-08-246-982A-6
; Sequence 6, Application US/08246982A
; Patent No. 5686288
; GENERAL INFORMATION:
; APPLICANT: MacDonald, Marcy E.
; APPLICANT: Ambrose, Christine M.
; APPLICANT: Duyao, Mabel P.
; APPLICANT: Guealla, James F.
; TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

US-08-453-265-6

Query Match 2.3%; Score 9; DB 1; Length 3144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
DB 1222 TTSKSSSLG 1230

RESULT 4

US-08-457-273B-42
; Sequence 42, Application US/08457273B
; Patent No. 5849995
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael
; APPLICANT: Lin, Biaoyang
; APPLICANT: Nasir, Jamal
; TITLE OF INVENTION: Mouse Model for Huntington's Disease and
; TITLE OF INVENTION: Related DNA Sequences
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 5849995th Carolina
; COUNTRY: US
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,273B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 3477-85A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-457-273B-42

QY 218 TTSKSSSLG 226
DB 1222 TTSKSSSLG 1230

RESULT 5

US-08-556-419-21
; Sequence 21, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
; APPLICANT: Ross, Christopher
; APPLICANT: Li, Xiao-Jiang
; APPLICANT: Li, Shi-Hua
; APPLICANT: Sharp, Alan
; APPLICANT: Lanahan, Anthony

US-08-453-265-6

Query Match 2.3%; Score 9; DB 1; Length 3144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
DB 1222 TTSKSSSLG 1230

RESULT 4

US-08-457-273B-42
; Sequence 42, Application US/08457273B
; Patent No. 5849995
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael
; APPLICANT: Lin, Biaoyang
; APPLICANT: Nasir, Jamal
; TITLE OF INVENTION: Mouse Model for Huntington's Disease and
; TITLE OF INVENTION: Related DNA Sequences
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 5849995th Carolina
; COUNTRY: US
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,273B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 3477-85A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-457-273B-42

QY 218 TTSKSSSLG 226
DB 1222 TTSKSSSLG 1230

RESULT 5

US-08-556-419-21
; Sequence 21, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
; APPLICANT: Ross, Christopher
; APPLICANT: Li, Xiao-Jiang
; APPLICANT: Li, Shi-Hua
; APPLICANT: Sharp, Alan
; APPLICANT: Lanahan, Anthony

US-08-453-265-6

Query Match 2.3%; Score 9; DB 1; Length 3144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
DB 1222 TTSKSSSLG 1230

RESULT 4

US-08-457-273B-42
; Sequence 42, Application US/08457273B
; Patent No. 5849995
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael
; APPLICANT: Lin, Biaoyang
; APPLICANT: Nasir, Jamal
; TITLE OF INVENTION: Mouse Model for Huntington's Disease and
; TITLE OF INVENTION: Related DNA Sequences
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 5849995th Carolina
; COUNTRY: US
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,273B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 3477-85A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3144 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-457-273B-42

QY 218 TTSKSSSLG 226
DB 1222 TTSKSSSLG 1230

RESULT 5

US-08-556-419-21
; Sequence 21, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
; APPLICANT: Ross, Christopher
; APPLICANT: Li, Xiao-Jiang
; APPLICANT: Li, Shi-Hua
; APPLICANT: Sharp, Alan
; APPLICANT: Lanahan, Anthony

APPLICANT: Worley, Paul
APPLICANT: Snyder, Solomon
TITLE OF INVENTION: Huntingtin-associated protein
FILE REFERENCE: 01107.52271
CURRENT APPLICATION NUMBER: US/08/556,419C
CURRENT FILING DATE: 1995-11-09
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 21
LENGTH: 3144
TYPE: PRT
ORGANISM: Homo sapiens
US-08-556-419-21

Query Match 2.3%; Score 9; DB 2; Length 3144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
Db 1222 TTSKSSSLG 1230

RESULT 6
US-09-041-886-15
Sequence 15, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
APPLICANT: Razizadeh, Sharrooz
TITLE OF INVENTION: Protoprotic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 3144 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-041-886-15

Query Match 2.3%; Score 9; DB 2; Length 3144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
Db 1222 TTSKSSSLG 1230

RESULT 7
US-09-538-092-1118
Sequence 1118, Application US/09538092
Patent No. 6753314
GENERAL INFORMATION:
APPLICANT: Giot, Loic
APPLICANT: Mansfield, Traci A.
TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
FILE REFERENCE: 15966-542
CURRENT APPLICATION NUMBER: US/09/538,092
CURRENT FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: 60/127,352
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: 60/178,965
PRIOR FILING DATE: 2000-02-01
NUMBER OF SEQ ID NOS: 1387
SOFTWARE: CurapatSeqformat Version 0.9
SEQ ID NO 1118
LENGTH: 3144
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (0)...(0)
OTHER INFORMATION: Polypeptide Accession Number P42858
US-09-538-092-1118

Query Match 2.3%; Score 9; DB 2; Length 3144;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
Db 1222 TTSKSSSLG 1230

RESULT 8
US-09-270-767-32008
Sequence 32008, Application US/09270767
Patent No. 6703491
GENERAL INFORMATION:
APPLICANT: Homburger et al.
TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
FILE REFERENCE: File Reference: 7326-094
CURRENT APPLICATION NUMBER: US/09/270,767
CURRENT FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 62517
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 32008
LENGTH: 136
TYPE: PRT
ORGANISM: Drosophila melanogaster
US-09-270-767-32008

Query Match 2.0%; Score 8; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 HLETVIPA 336
Db 72 HLETVIPA 79

RESULT 9
US-09-826-509-491
Sequence 491, Application US/09826509
Patent No. 6806054
GENERAL INFORMATION:
APPLICANT: Lehmann-Bruinsma, Karin
APPLICANT: Lin, I-Lin
TITLE OF INVENTION: No. 6806054-Endogenous, Constitutively Activated Known G
TITLE OF INVENTION: Protein-Coupled Receptors

FILE REFERENCE: AREN-207
CURRENT APPLICATION NUMBER: US/09/826,509
CURRENT FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/195,747
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: 09/170,496
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 589
SOFTWARE: Patentin Version 2.1
SEQ ID NO 491
LENGTH: 400
TYPE: PRT
ORGANISM: Homo sapiens
US-09-826-509-491

Query Match 2.0%; Score 8; DB 2; Length 400;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133

RESULT 10
US-07-626-618A-20
Sequence 20, Application US/07626618A
Patent No. 5422265
GENERAL INFORMATION:
APPLICANT: Van Tol, Hubert H.M.
APPLICANT: Civeilli, Olivier
TITLE OF INVENTION: A No. 5422265el Human Dopamine Receptor and Uses
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/626,618A
FILING DATE: 7 DEC 1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: NO. 5422265nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 90,1092
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 810-221-8317
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO

US-07-626-618A-20
Query Match 2.0%; Score 8; DB 1; Length 444;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133

RESULT 11
US-08-333-977-20
Sequence 20, Application US/08333977
Patent No. 5594108
GENERAL INFORMATION:
APPLICANT: Van Tol, Hubert H.M.
APPLICANT: Civeilli, Olivier
TITLE OF INVENTION: A No. 5594108el Human Dopamine Receptor and Uses
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/333,977
FILING DATE: 03-NOV-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/626,618
FILING DATE: 7 DEC 1990
ATTORNEY/AGENT INFORMATION:
NAME: NO. 5594108nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 90,1092
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 810-221-8317
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 444 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO

Query Match 2.0%; Score 8; DB 1; Length 444;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
Db 126 IDRYTAVV 133

RESULT 12
US-07-781-254A-2
Sequence 2, Application US/07781254A
Patent No. 5407823
GENERAL INFORMATION:
APPLICANT: Sokoloff, Pierre
APPLICANT: Martress, Marie-Pascale
APPLICANT: Schwartz, Jean-Charles
APPLICANT: Giros, Bruno
TITLE OF INVENTION: Polypeptides Having a Dopaminergic Receptor
TITLE OF INVENTION: Activity, Nucleic Acids Coding for These Polypeptides and the
TITLE OF INVENTION: Use of These Polypeptides for the Screening of Substances
TITLE OF INVENTION: Active on These Polypeptides
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchand & Gould
STREET: 3100 No. 5407823west Center

;; CITY: Minneapolis
;; STATE: MN
;; COUNTRY: USA
;; ZIP: 55402
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/781,254A
;; FILING DATE: 31-DEC-1991
;; CLASSIFICATION: 530
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hillson, Randall A.
;; REGISTRATION NUMBER: 31,939
;; REFERENCE/DOCKET NUMBER: 8076.51-WOUS
;;
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 612-332-5300
;; TELEFAX: 612-332-9081
;;
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 446 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; IMMEDIATE SOURCE:
;; CLONE: Variant D-3 dopaminergic receptor
;;
US-07-781-254A-2

Query Match 2.0%; Score 8; DB 1; Length 446;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
|||
Db 126 IDRYTAVV 133

RESULT 13
US-07-781-254A-3
; Sequence 3, Application US/07781254A
; Patent No. 5407823
; GENERAL INFORMATION:
; APPLICANT: Sokoloff, Pierre
; APPLICANT: Martress, Marie-Pascale
; APPLICANT: Schwartz, Jean-Charles
; APPLICANT: Giros, Bruno
; TITLE OF INVENTION: Polypeptides Having a Dopaminergic Receptor
; TITLE OF INVENTION: Activity, Nucleic Acids Coding for These Polypeptides and the
; TITLE OF INVENTION: Use of These Polypeptides for the Screening of Substances
; TITLE OF INVENTION: Active on These Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5407823west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/781,254A
; FILING DATE: 31-DEC-1991
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Hillson, Randall A.
; REGISTRATION NUMBER: 31,939
; REFERENCE/DOCKET NUMBER: 8076.51-WOUS

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 612-332-5300
;; TELEFAX: 612-332-9081
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 446 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; IMMEDIATE SOURCE:
;; CLONE: D-3 dopaminergic receptor
;;
US-07-781-254A-3

Query Match 2.0%; Score 8; DB 1; Length 446;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 307 IDRYTAVV 314
|||
Db 126 IDRYTAVV 133

RESULT 14
US-09-252-991A-18637
; Sequence 18637, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18637
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18637

Query Match 2.0%; Score 8; DB 2; Length 546;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GAGGSAGA 217
|||
Db 520 GAGGSAGA 527

RESULT 15
US-09-949-002-426
; Sequence 426, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 426
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Human

US-09-949-002-426

Query Match 2.0%; Score 8; DB 2; Length 562;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 210 GAGGSAGA 217
|||
Db 96 GAGGSAGA 103
|||

Search completed: November 23, 2005, 16:50:48
Job time : 33.9158 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:44:35 ; Search time 110.717 Seconds
(without alignments)
1475.577 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 391
Sequence: 1 MGKSDFIQPKAIANRIKSG.....GPLKGRVGGIOVEDISKLA 391

Scoring table: OLIGO
Gapop 60.0 , Capext 60.0

Searched: 1867569 seqs, 417829326 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap:*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap:*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap:*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pap:*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pap:*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	179	45.8	393	4	US-10-408-765A-1465
2	35	9.0	120	4	US-10-767-701-57151, A
3	35	9.0	138	4	US-10-437-963-191916
4	35	9.0	423	5	US-10-739-930-8499
5	35	9.0	424	4	US-10-425-115-341788
6	35	9.0	437	4	US-10-425-114-48183
7	35	9.0	437	4	US-10-425-114-58982
8	30	7.7	176	4	US-10-424-599-168326
9	29	7.4	84	4	US-10-106-698-5464
10	15	3.8	390	6	US-11-097-143-41040
11	9	2.3	335	5	US-10-732-923-11274
12	9	2.3	535	5	US-10-739-930-10661
13	9	2.3	539	5	US-10-732-923-11269
14	9	2.3	1543	3	US-09-904-987-7
15	9	2.3	3141	4	US-10-408-765A-2250
16	9	2.3	3144	4	US-10-419-597-1
17	9	2.3	3223	4	US-10-276-774-1777
18	9	2.3	3223	4	US-10-276-774-1840
19	8	2.0	8	4	US-10-022-066-38
20	8	2.0	8	4	US-10-022-066-610
21	8	2.0	10	3	US-09-572-404B-3236
22	8	2.0	53	4	US-10-425-115-296022
23	8	2.0	143	4	US-10-437-963-176007
24	8	2.0	164	4	US-10-425-115-287597
25	8	2.0	183	6	US-11-097-143-27039
26	8	2.0	192	6	US-11-097-143-2925
27	8	2.0	198	3	US-09-794-257-16

28	8	2.0	198	3	US-09-945-173-5	Sequence 5, Appl1
29	8	2.0	198	3	US-09-972-529-4	Sequence 4, Appl1
30	8	2.0	198	4	US-10-400-991-70	Sequence 70, Appl1
31	8	2.0	222	4	US-10-180-375-110	Sequence 126, App
32	8	2.0	222	4	US-10-183-687-126	Sequence 126, App
33	8	2.0	224	4	US-10-424-599-202515	Sequence 202515, App
34	8	2.0	224	5	US-10-739-930-9455	Sequence 9455, App
35	8	2.0	229	4	US-10-180-375-112	Sequence 112, App
36	8	2.0	229	4	US-10-183-687-128	Sequence 128, App
37	8	2.0	229	4	US-10-424-599-178340	Sequence 178340, App
38	8	2.0	362	4	US-10-424-599-247365	Sequence 247365, App
39	8	2.0	400	3	US-09-826-509-431	Sequence 491, App
40	8	2.0	400	4	US-10-225-567A-104	Sequence 104, App
41	8	2.0	400	4	US-10-292-798-1244	Sequence 1244, App
42	8	2.0	400	5	US-10-925-095-491	Sequence 491, App
43	8	2.0	438	4	US-10-425-115-214055	Sequence 214055, App
44	8	2.0	461	4	US-10-282-122A-64750	Sequence 64750, A
45	8	2.0	494	4	US-10-282-122A-52864	Sequence 52864, A

ALIGNMENTS

RESULT 1
US-10-408-765A-1465
; Sequence 1465, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1465
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1465

Query Match	45.8%	Score 179;	DB 4;	Length 393;
Best Local Similarity	100.0%;	Pred. No. 4.4e-159;		
Matches 179;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	9	PKAIANRIKSGKLGKLRWTCQMCQKCRDENGPKCHCMESHOROLLASENPQQPMDFY 68		
Db	9	PKAIANRIKSGKLGKLRWTCQMCQKCRDENGPKCHCMESHOROLLASENPQQPMDFY 68		
QY	69	SEEFNDFLELLRRRGTGKRVHNNIVYNEIYSHREIHNNATOWETLDTFTKWLGRGLC 128		
Db	69	SEEFNDFLELLRRRGTGKRVHNNIVYNEIYSHREIHNNATOWETLDTFTKWLGRGLC 128		
QY	129	KVDETQKGMWYIQVIDRDPETIRQLELEKKKKODDDEKTAKFIEQVRRGLGKQEQ 187		
Db	129	KVDETQKGMWYIQVIDRDPETIRQLELEKKKKODDDEKTAKFIEQVRRGLGKQEQ 187		

RESULT 2
US-10-767-701-57151
; Sequence 57151, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53353)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 57151
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30563818.pcp
US-10-767-701-57151

Query Match 9.0%; Score 35; DB 4; Length 120;
Best Local Similarity 100.0%; Pred. No. 2.4e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53
DB 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53

RESULT 3

US-10-437-963-191916
; Sequence 191916, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 191916
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1.pcp
US-10-437-963-191916

Query Match 9.0%; Score 35; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 2.8e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53
DB 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53

RESULT 4

US-10-739-930-8499
; Sequence 8499, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 8499
; LENGTH: 423

; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C7583_2.p
US-10-739-930-8499

Query Match 9.0%; Score 35; DB 5; Length 423;
Best Local Similarity 100.0%; Pred. No. 7.4e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53
DB 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53

RESULT 5

US-10-425-115-341788
; Sequence 341788, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 341788
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_74879C.1.pcp
US-10-425-115-341788

Query Match 9.0%; Score 35; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 7.4e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53
DB 19 KGLQKLRLWYCMQCKQCRDENGFKCHCMSESHQRQ 53

RESULT 6

US-10-425-114-48183
; Sequence 48183, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53131)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48183
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI.pcp
US-10-425-114-48183

Query Match 9.0%; Score 35; DB 4; Length 437;

```
Best Local Similarity 100.0%; Pred. No. 7.6e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQRQ 53
DB 32 KGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQRQ 66

RESULT 7
US-10-425-114-58982
; Sequence 58982, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 58982
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700238628_FLI.pep
US-10-425-114-58982

Query Match 9.0%; Score 35; DB 4; Length 437;
Best Local Similarity 100.0%; Pred. No. 7.6e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQRQ 53
DB 32 KGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQRQ 66

RESULT 8
US-10-424-599-168326
; Sequence 168326, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 168326
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1.pep
US-10-424-599-168326

Query Match 7.7%; Score 30; DB 4; Length 176;
Best Local Similarity 100.0%; Pred. No. 1.7e-19;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYWCQMCQKQCRDENGFKCHCMSE 48
DB 19 KGLQKLRYWCQMCQKQCRDENGFKCHCMSE 48
```

```
RESULT 9
US-10-106-698-5464
; Sequence 5464, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5464
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5464

Query Match 7.4%; Score 29; DB 4; Length 84;
Best Local Similarity 100.0%; Pred. No. 7.6e-19;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 363 FSATIVTGTPLKGRVRVEGIQYEDISKLA 391
DB 56 FSATIVTGTPLKGRVRVEGIQYEDISKLA 84

RESULT 10
US-11-097-143-41040
; Sequence 41040, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41040
; LENGTH: 390
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-41040

Query Match 3.8%; Score 15; DB 6; Length 390;
Best Local Similarity 100.0%; Pred. No. 4.1e-05;
```

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 KSKGLQKLRYQCMC 31
|||||
Db 17 KSKGLQKLRYQCMC 31

RESULT 11
US-10-732-923-11274
; Sequence 11274, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 11274
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-10-732-923-11274

Query Match 2.3%; Score 9; DB 5; Length 335;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ELSRENEEE 201
|||||
Db 318 ELSRENEEE 326

RESULT 12
US-10-739-930-10661
; Sequence 10661, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 10661
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; OTHER INFORMATION: Clone ID: TRIAE-23APR03-C568_2.p
US-10-739-930-10661

Query Match 2.3%; Score 9; DB 5; Length 535;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ELSRENEEE 201
|||||
Db 518 ELSRENEEE 526

RESULT 13
US-10-732-923-11269
; Sequence 11269, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923

; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 11269
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(539)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-732-923-11269

Query Match 2.3%; Score 9; DB 5; Length 539;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 ELSRENEEE 201
|||||
Db 521 ELSRENEEE 529

RESULT 14
US-09-904-987-7
; Sequence 7, Application US/09904987
; Patent No. US20020037908A1
; GENERAL INFORMATION:
; APPLICANT: No. US20020037908A1actyl, Inc.
; TITLE OF INVENTION: Methods and Compositions for Controlling Pathological and Prepathe
; TITLE OF INVENTION: Protein Assembly or Aggregation
; FILE REFERENCE: 42108/26146
; CURRENT APPLICATION NUMBER: US/09/904,987
; CURRENT FILING DATE: 2001-07-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1543
; TYPE: PRT
; ORGANISM: homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NCBI ENTREZ / XP_003405
; DATABASE ENTRY DATE: 2001-04-16
; RELEVANT RESIDUES: (1)..(1543)
US-09-904-987-7

Query Match 2.3%; Score 9; DB 3; Length 1543;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 TTSKSSSLG 226
|||||
Db 1219 TTSKSSSLG 1227

RESULT 15
US-10-408-765A-2250
; Sequence 2250, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2250
; LENGTH: 3141
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2250

Query Match 2.3%; Score 9; DB 4; Length 3141;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 TTSKSSSLG 226
|||
Db 1219 TTSKSSSLG 1227

Search completed: November 23, 2005, 16:55:48
Job time : 111.717 secs

This Page Blank (uspic)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:40:00 ; Search time 4.98725 Seconds
(without alignments)
237.628 Million cell updates/sec

Title: US-09-555-529-25
Perfect score: 391
Sequence: 1 MKKSFSLSPKATANKSG.....GPLKGRVGIQYEDISKLA 391

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 17545 seqs, 3030971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 17545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA New.*
1: /cgn2_6/prodata/2/pubaa/US10_NEW_PUB.pep.*
2: /cgn2_6/prodata/2/pubaa/US06_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubaa/US07_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubaa/US08_NEW_PUB.pep.*
5: /cgn2_6/prodata/2/pubaa/US09_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubaa/PCT_NEW_PUB.pep.*
7: /cgn2_6/prodata/2/pubaa/US11_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	9	2.3	3144	7 US-11-055-035-1	Sequence 1, Appli
2	7	1.8	229	7 US-11-014-629-1	Sequence 1, Appli
3	7	1.8	318	1 US-10-802-796-727	Sequence 727, App
4	7	1.8	334	1 US-10-802-796-728	Sequence 728, App
5	6	1.5	112	1 US-10-793-626-2114	Sequence 2114, Ap
6	6	1.5	149	1 US-10-986-501-231	Sequence 231, App
7	6	1.5	204	1 US-10-793-626-2394	Sequence 2394, Ap
8	6	1.5	215	1 US-10-821-234-1443	Sequence 1443, Ap
9	6	1.5	227	1 US-10-793-626-2642	Sequence 2642, Ap
10	6	1.5	286	7 US-11-137-465-48	Sequence 48, Appl
11	6	1.5	339	1 US-10-821-234-1507	Sequence 1507, Ap
12	6	1.5	346	1 US-10-967-648A-10	Sequence 10, Appl
13	6	1.5	370	1 US-10-821-234-1105	Sequence 1105, Ap
14	6	1.5	389	7 US-11-012-762-72	Sequence 72, Appl
15	6	1.5	391	1 US-10-996-628-2	Sequence 2, Appli
16	6	1.5	398	7 US-11-012-762-74	Sequence 74, Appl
17	6	1.5	441	1 US-10-510-386-162	Sequence 162, App
18	6	1.5	447	7 US-10-821-234-1584	Sequence 1584, Ap
19	6	1.5	458	7 US-11-077-550-114	Sequence 114, App
20	6	1.5	460	1 US-10-858-730-63	Sequence 63, Appl
21	6	1.5	467	1 US-10-510-386-56	Sequence 56, Appl
22	6	1.5	468	7 US-11-076-187-2	Sequence 2, Appli
23	6	1.5	548	7 US-11-077-550-24	Sequence 24, Appl
24	6	1.5	600	1 US-10-131-826A-462	Sequence 462, App
25	6	1.5	633	1 US-10-821-234-1146	Sequence 1146, Ap

26	6	1.5	724	1 US-10-793-626-968	Sequence 968, App
27	6	1.5	749	1 US-10-793-626-138	Sequence 138, App
28	6	1.5	759	1 US-10-858-730-75	Sequence 75, Appl
29	6	1.5	760	1 US-10-858-730-76	Sequence 76, Appl
30	6	1.5	853	1 US-10-821-234-1110	Sequence 1110, Ap
31	6	1.5	858	7 US-11-077-550-22	Sequence 22, Appl
32	6	1.5	860	7 US-11-077-550-175	Sequence 175, App
33	6	1.5	862	7 US-11-077-550-94	Sequence 94, Appl
34	6	1.5	862	7 US-11-077-550-171	Sequence 171, App
35	6	1.5	862	7 US-11-077-550-173	Sequence 173, App
36	6	1.5	864	7 US-11-077-550-102	Sequence 102, App
37	6	1.5	865	7 US-11-077-550-100	Sequence 100, App
38	6	1.5	866	7 US-11-077-550-88	Sequence 88, Appl
39	6	1.5	866	7 US-11-077-550-104	Sequence 104, App
40	6	1.5	867	7 US-11-077-550-80	Sequence 80, Appl
41	6	1.5	867	7 US-11-077-550-96	Sequence 96, Appl
42	6	1.5	867	7 US-11-077-550-98	Sequence 98, Appl
43	6	1.5	870	7 US-11-077-550-92	Sequence 92, Appl
44	6	1.5	871	7 US-11-077-550-84	Sequence 84, Appl
45	6	1.5	871	7 US-11-077-550-86	Sequence 86, Appl

ALIGNMENTS

RESULT 1
US-11-055-035-1
; Sequence 1, Application US/11055035
; Publication No. US20050256072A1
; GENERAL INFORMATION:
; APPLICANT: ARONIN, NEIL
; APPLICANT: ZAMORE, PHILLIP D.
; APPLICANT: BRODERICK, JENNIFER
; TITLE OF INVENTION: DUAL FUNCTIONAL OLIGONUCLEOTIDES FOR USE IN REPRESSING
; FILE REFERENCE: UMY-095
; CURRENT APPLICATION NUMBER: US/11/055,035
; CURRENT FILING DATE: 2005-02-09
; PRIOR APPLICATION NUMBER: 60/543,467
; PRIOR FILING DATE: 2004-02-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patent In Ver. 3.3
; SEQ ID NO 1
; LENGTH: 3144
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-055-035-1

Query Match	2.3%	Score 9;	DB 7;	Length 3144;
Best Local Similarity	100.0%	Pred. No. 0.89;		
Matches	9;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
QY	218	TTSKSSSLG 226		
Db	1222	TTSKSSSLG 1230		
RESULT 2				
US-11-014-629-1				
; Sequence 1, Application US/11014629				
; Publication No. US20050244376A1				
; GENERAL INFORMATION:				
; APPLICANT: Hall, Frederick L.				
; APPLICANT: Gordon, Erlinda M.				
; TITLE OF INVENTION: TARGETED GENE DELIVERY TO EXTRACELLULAR MATRIX				
; FILE REFERENCE: 30863-704.302				
; CURRENT APPLICATION NUMBER: US/11/014,629				
; CURRENT FILING DATE: 2004-12-15				
; PRIOR APPLICATION NUMBER: US 08/837,223				
; PRIOR FILING DATE: 1997-04-10				
; PRIOR APPLICATION NUMBER: US 09/904,923				
; PRIOR FILING DATE: 2001-07-13				
; NUMBER OF SEQ ID NOS: 8				

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 229
; TYPE: PRT
; ORGANISM: moloney murine leukemia virus
US-11-014-629-1

Query Match
Best Local Similarity 1.8%; Score 7; DB 7; Length 229;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 RLKLDQT 328
Db 102 RLKLDQT 108

RESULT 3
US-10-802-796-727
; Sequence 727, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 727
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Mycobacterium sp.
US-10-802-796-727

Query Match
Best Local Similarity 1.8%; Score 7; DB 1; Length 318;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GAGGSAG 216
Db 202 GAGGSAG 208

RESULT 4
US-10-802-796-728
; Sequence 728, Application US/10802796
; Publication No. US20050250104A1
; GENERAL INFORMATION:
; APPLICANT: COLE, STEWART
; APPLICANT: BUCHRIESER-BROSCH, ROLAND
; APPLICANT: GORDON, STEPHEN
; APPLICANT: BILLAULT, ALAIN
; TITLE OF INVENTION: A METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST
; TITLE OF INVENTION: FROM THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED
; TITLE OF INVENTION: DNA LIBRARY. APPLICATION TO THE DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIA.
; FILE REFERENCE: 05394.0011-00000
; CURRENT APPLICATION NUMBER: US/10/802,796
; CURRENT FILING DATE: 2004-03-18
; PRIOR APPLICATION NUMBER: US/09/673,476
; PRIOR FILING DATE: 2002-03-29
```

```
; PRIOR APPLICATION NUMBER: PCT/IB99/00740
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.2
; SEQ ID NO 728
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Mycobacterium sp.
US-10-802-796-728

Query Match
Best Local Similarity 1.8%; Score 7; DB 1; Length 334;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 210 GAGGSAG 216
Db 173 GAGGSAG 179

RESULT 5
US-10-793-626-2114
; Sequence 2114, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2114
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (112)
; OTHER INFORMATION: variable amino acid
US-10-793-626-2114

Query Match
Best Local Similarity 1.5%; Score 6; DB 1; Length 112;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 QRQLLL 56
Db 35 QRQLLL 40

RESULT 6
US-10-986-501-231
; Sequence 231, Application US/10986501
; Publication No. US20050244845A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P2C1
; CURRENT APPLICATION NUMBER: US/10/986,501
; CURRENT FILING DATE: 2004-11-12
; PRIOR APPLICATION NUMBER: US/10/621,363
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: 09/969,730
; PRIOR FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: 09/774,639
; PRIOR FILING DATE: 2001-02-01
```

```
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; PRIOR FILING DATE: 1997-08-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 231
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-986-501-231

Query Match      1.5%; Score 6; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 LEKKK 160
Db 141 LEKKK 146

RESULT 7
US-10-793-626-2394
; Sequence 2394, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2394
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; US-10-793-626-2394

Query Match      1.5%; Score 6; DB 1; Length 204;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 QROLLL 56
Db 35 QROLLL 40

RESULT 8
US-10-821-234-1443
; Sequence 1443, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom

; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1443
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-821-234-1443

Query Match      1.5%; Score 6; DB 1; Length 215;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 299 KKGVK 304
Db 172 KKGVK 177

RESULT 9
US-10-793-626-2642
; Sequence 2642, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2642
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; US-10-793-626-2642

Query Match      1.5%; Score 6; DB 1; Length 227;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 353 GTLESI 358
Db 57 GTLESI 62

RESULT 10
US-11-137-465-48
; Sequence 48, Application US/11137465
; Publication No. US20050255556A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdoch, Paul R.
; APPLICANT: Rizvi, Safia, K.
; APPLICANT: Smith, Randall, F.
; APPLICANT: Xiang, Zhaoying
; APPLICANT: Kabnick, Karen
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50018
; CURRENT APPLICATION NUMBER: US/11/137,465
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US/10/239,663
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/09226
```

```
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/192,158
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/192,668
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/200,166
; PRIOR FILING DATE: 2000-04-27
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 48
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-137-465-48

Query Match      1.5%; Score 6; DB 7; Length 286;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 FTLSR 196
Db 172 FTLSR 177

RESULT 11
US-10-821-234-1507
; Sequence 1507, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1507
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1507

Query Match      1.5%; Score 6; DB 1; Length 339;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 HLETVI 334
Db 94 HLETVI 99

RESULT 12
US-10-967-648A-10
; Sequence 10, Application US/10967648A
; Publication No. US20050245473A1
; GENERAL INFORMATION:
; APPLICANT: Saunders, Nicholas A
; TITLE OF INVENTION: Differentiation- and/or proliferation-modulating agents and uses
; FILE REFERENCE: 12493972
; CURRENT APPLICATION NUMBER: US/10/967,648A
; CURRENT FILING DATE: 2004-10-15
; PRIOR APPLICATION NUMBER: USSN 60/512010
; PRIOR FILING DATE: 2003-10-16
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 10
; LENGTH: 346
```

```
; TYPE: PRT
; ORGANISM: Human
US-10-967-648A-10

Query Match      1.5%; Score 6; DB 1; Length 346;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 304 KEVIDR 309
Db 124 KEVIDR 129

RESULT 13
US-10-821-234-1105
; Sequence 1105, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1105
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1105

Query Match      1.5%; Score 6; DB 1; Length 370;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 GGSAGA 217
Db 78 GGSAGA 83

RESULT 14
US-11-012-762-72
; Sequence 72, Application US/11012762
; Publication No. US20050244815A1
; GENERAL INFORMATION:
; APPLICANT: Georgia State University Research Foundation, Inc.
; TITLE OF INVENTION: Compositions and Methods for Viral Resistance Genes
; FILE REFERENCE: GSU1.PCT
; CURRENT APPLICATION NUMBER: US/11/012,762
; CURRENT FILING DATE: 2004-12-15
; PRIOR APPLICATION NUMBER: PCT/US03/19300
; PRIOR FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US 60/390,046
; PRIOR FILING DATE: 2002-06-19
; NUMBER OF SEQ ID NOS: 133
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 72
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Bos taurus
US-11-012-762-72

Query Match      1.5%; Score 6; DB 7; Length 389;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 IRROLE 154
Db 149 IRROLE 154
```

Wed Nov 30 09:40:47 2005

Db 101 IRRQLE 106

RESULT 15
US-10-996-628-2
; Sequence 2, Application US/10996628
; Publication No. US20050244415A1
; GENERAL INFORMATION:
; APPLICANT: Hiroshi Ohnishi et al.
; TITLE OF INVENTION: Pharmaceutical composition and method for activating macrophage
; TITLE OF INVENTION: same
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/996.628
; CURRENT FILING DATE: 2004-11-23
; PRIOR APPLICATION NUMBER: JP2004-133599
; PRIOR FILING DATE: 2004-04-28
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-996-628-2

Query Match 1.5%; Score 6; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 210 GAGGSA 215
|||
Db 43 GAGGSA 48

Search completed: November 23, 2005, 16:52:01
Job time : 5.98725 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:06:33 ; Search time 130.833 Seconds
(without alignments)
1319.820 Million cell updates/sec

Title: US-09-555-529-26

Perfect score: 2037

Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPKLGRVGEIQVEDISKLA 393

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1980s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2037	100.0	393	2	AAY23620 A human k
2	2037	100.0	393	7	ADJ69659 Human hea
3	1887	92.6	391	2	AAY23619 Murine ki
4	1465	71.9	293	2	AAY23618 A huamn t
5	1365	67.0	291	2	AAY23617 Mouse tru
6	1243	61.0	390	8	ADP22452 Sea-squir
7	1059.5	52.0	390	4	ABE71416 Drosophil
8	978	48.0	411	3	AAG42576 Arabidops
9	934	45.9	424	3	AAAB03063 Maize KIN
10	934	45.9	437	8	ADX96318 Plant ful
11	934	45.9	437	8	ADX78817 Plant ful
12	909.5	44.6	423	3	AB03064 Maize KIN
13	909.5	44.6	423	3	AB03065 Maize KIN
14	909.5	44.6	423	8	ADT58422 Plant pol
15	829	40.7	382	3	AAG42577 Arabidops
16	726	35.6	366	3	AAG42578 Arabidops
17	615	30.2	340	3	AAG46711 Arabidops
18	615	30.2	343	3	AAG46710 Arabidops
19	599	29.4	302	3	AAG46712 Arabidops
20	370	18.2	84	4	AAG74690 Human col
21	176.5	8.7	82	3	AAG19349 Arabidops
22	173	8.5	81	3	AAG19350 Arabidops
23	165	8.1	76	3	AAG19351 Arabidops
24	155	7.6	76	3	AAG46729 Arabidops

25	150.5	7.4	278	8	ADX77206	Adx77206 Plant ful
26	145.5	7.1	1114	7	ADJ70313	Adj70313 Human hea
27	145.5	7.1	1213	4	AAW40016	Aam40016 Human pol
28	145.5	7.1	1213	6	ABP97031	Abp97031 Human L-F
29	144.5	7.1	1212	6	ABP97029	Abp97029 Rat L-FIL
30	144	7.1	375	8	AG21445	Aag21445 Arabidops
31	140	7.1	375	8	ADT56172	Adt56172 Plant pol
32	140	6.9	375	3	AAG43126	Aag43126 Arabidops
33	140	6.9	450	3	AAG50080	Aag50080 Arabidops
34	139	6.8	348	7	AD140582	Adi40582 Human pur
35	138.5	6.8	517	2	AAR22904	Aar22904 1-Caldeem
36	138	6.8	805	8	ADN17499	Adn17499 Bacterial
37	137	6.7	323	3	AAG15386	Aag15386 Arabidops
38	136	6.7	347	4	AAW41227	Aam41227 Human pol
39	136	6.7	508	7	ADD01217	Add01217 Human nuc
40	136	6.7	508	8	ADN06062	Adn06062 Antipsori
41	136	6.7	624	5	AAE15257	Aae15257 Human RNA
42	136	6.7	624	6	ABU62546	Abu62546 Human PNI
43	135.5	6.7	237	3	AAG15388	Aag15388 Arabidops
44	135.5	6.7	246	3	AAG15387	Aag15387 Arabidops
45	135.5	6.7	512	7	ADM05395	Adm05395 Human pro

ALIGNMENTS

RESULT 1
AAY23620
ID AAY23620 standard; protein; 393 AA.
XX
AC AAY23620;
XX
DT 07-SEP-1999 (first entry)
XX
DE A human kin17 protein.
XX
KW Human; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme.
XX
OS Homo sapiens.
PN FR2772046-A1.
PD 11-JUN-1999.
PF 09-DEC-1997; 97FR-00015536.
PR 09-DEC-1997; 97FR-00015536.
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
WPI; 1999-359999/31.
New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.
Claim 19; Page 37-38; 69pp; French.
The present sequence represents a human kin17 protein. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation

Sequence 393 AA;

PT	treating diseases associated with altered mitochondrial function,				
PT	comprises detecting a modified polypeptide in a sample and correlating				
PT	with the disease.				
XX					
PS	Claim 1; SEQ ID NO 1465; 180pp; English.				
XX					
CC	This invention relates to novel mitochondrial targets that can be used				
CC	for therapeutic intervention in treating a disease associated with				
CC	altered mitochondrial function. Specifically, it refers to a method for				
CC	identifying proteins of the human heart mitochondrial proteome that are				
CC	useful for drug screening assays, as well as therapeutic targets. The				
CC	present invention describes a method for identifying such proteins that				
CC	can be used in the treatment of various diseases associated with altered				
CC	mitochondrial function including diabetes mellitus, Huntington's disease,				
CC	osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial				
CC	encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy				
CC	ragged red fibre syndrome (MERRF) or cancer. Accordingly, these				
CC	compositions have neuroprotective, nontropic, antidiabetic,				
CC	anticonvulsant, antiarthritic, osteopathic, ophthalmological and				
CC	cytostatic activities. This polypeptide sequence is a human heart				
CC	mitochondrial protein of the invention.				
XX					
SQ	Sequence 393 AA;				
Query Match		100.0%;	Score 2037;	DB 7;	Length 393;
Best Local Similarity		100.0%;	Pred. No. 2.1e-168;		
Matches 393;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKSDFLTPKAIANRIKSKGLQKLRWYCMQCQCRDENGFKCHCMSESHORQLLASEN	60		
DB	1	MKSDFLTPKAIANRIKSKGLQKLRWYCMQCQCRDENGFKCHCMSESHORQLLASEN	60		
QY	61	PQOFMDYFSEEFNDLELLRRRFGTKRVHNNIVYNEIYISHREHIMNATQWETLTDFTK	120		
DB	61	PQOFMDYFSEEFNDLELLRRRFGTKRVHNNIVYNEIYISHREHIMNATQWETLTDFTK	120		
QY	121	WLREGCLKVDKTPKGYIYIDRDPETIRROLELEKKKKQDLDDDEKTAFTIEEQVRRG	180		
DB	121	WLREGCLKVDKTPKGYIYIDRDPETIRROLELEKKKKQDLDDDEKTAFTIEEQVRRG	180		
QY	181	LEGKEQEVPTFTLSRENDEKVTFTNLSKGACSSSGATSSKSTLGPALKTTIGSSASVK	240		
DB	181	LEGKEQEVPTFTLSRENDEKVTFTNLSKGACSSSGATSSKSTLGPALKTTIGSSASVK	240		
QY	241	RKSSQSSTQSEKKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVKIITKKLGEKYHK	300		
DB	241	RKSSQSSTQSEKKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVKIITKKLGEKYHK	300		
QY	301	KKAIKVEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAGKRILVLNGGYRGNEGTLTLESI	360		
DB	301	KKAIKVEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAGKRILVLNGGYRGNEGTLTLESI	360		
QY	361	NEKTFSAIVTETGPKLGRVREGIQVEDISKLA	393		
DB	361	NEKTFSAIVTETGPKLGRVREGIQVEDISKLA	393		
RESULT 3					
AAJ23619					
ID	AAJ23619	standard; protein; 391 AA.			
XX	XX	AAJ23619;			
AC	AAJ23619;				
DT	07-SEP-1999	(first entry)			
XX	XX	Murine kin17 protein.			
DE	DE	Mouse; kin17 protein; cell proliferation; fertility;			
XX	XX	KW hyperproliferative disease; protein interaction; curved DNA;			
KW	KW	HIV replication; HIV integration; repair enzyme.			
XX	XX	Mus sp.			
OS	OS				
XX	XX				


```

PN FR2772046-A1.
XX
XX
PD 11-JUN-1999.
XX
XX PF 09-DEC-1997; 97FR-00015536.
XX
XX PR 09-DEC-1997; 97FR-00015536.
XX
XX (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX
XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
XX
XX New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.
XX
XX Claim 19; Page 36-37; 69pp; French.
XX
XX The present sequence represents a murine kin17 protein. The mammalian
XX kin17 protein is useful for preparing a medicament for controlling cell
XX proliferation or for controlling fertility. The medicaments can also be
XX used to treat hyperproliferative diseases. Fragments between amino acids
XX 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
XX kin17 protein are useful for regulating the interaction between proteins
XX and curved DNA. The fragment can be used to block replication of HIV or
XX its integration into the human genome or to target repair enzymes to
XX curved DNA sites. Expression vectors for kin17 can be used for
XX controlling cell proliferation
XX
XX Sequence 391 AA;

Query Match 92.6%; Score 1887; DB 2; Length 391;
Best Local Similarity 92.1%; Pred. No. 2.5e-155;
Matches 362; Conservative 14; Mismatches 15; Indels 2; Gaps 1;

Qy 1 MGKSDFLTTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGKSDFLTTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60

Qy 61 PQQFMDYFSEEFNDLFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDFTK 120
Db 61 PQQFMDYFSEEFNDLFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDFTK 120

Qy 121 WLGRGLCKVDETPKGWYIQVIDRDPETIRROLEKLEKKKQDLDDDEKTAFFIEEQVRG 180
Db 121 WLGRGLCKVDETPKGWYIQVIDRDPETIRROLEKLEKKKQDLDDDEKTAFFIEEQVRG 180

Qy 181 LEGKEQVPTFTELSENDEKVTFNLSKGACSSSGATSSKSTLGPALKTIGSSASVK 240
Db 181 LEGKEQVPTFTELSENDEKVTFNLSKGACSSSGATSSKSTLGPALKTIGSSASVK 240

Qy 241 RKSSQSSQSKKKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVKIIITKLGKGYHK 300
Db 241 RKSSQSSQSKKKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVKIIITKLGKGYHK 298

Qy 301 KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAPGRILVLNGVGRNGEGTLESI 360
Db 301 KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAPGRILVLNGVGRNGEGTLESI 358

Qy 361 NEKTFSAITIVETGPKGRRVEGIOYEDISKLA 393
Db 359 NEKAFSATIVETGPKGRRVEGIOYEDISKLA 391

RESULT 4
AY23618
ID AY23618 standard; protein; 293 AA.
XX
XX AY23618;
AC
XX
XX DT 07-SEP-1999 (first entry)
XX

```

A huamn truncated kin17 protein.

kin17 protein; cell proliferation; fertility; hyperproliferative disease; protein interaction; curved DNA; HIV replication; HIV integration; repair enzyme.

Homo sapiens.

FR2772046-A1.

11-JUN-1999.

09-DEC-1997; 97FR-00015536.

09-DEC-1997; 97FR-00015536.

(COMS) COMMISSARIAT ENERGIE ATOMIQUE.

Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF; WPI; 1999-359999/31.

New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.

Claim 15; Page 34-35; 69pp; French.

The present sequence represents a truncated human kin17 protein with amino acids 129-228 deleted. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation

Sequence 293 AA;

Query Match 71.9%; Score 1465; DB 2; Length 293;
Best Local Similarity 74.6%; Pred. No. 9e-119;
Matches 293; Conservative 0; Mismatches 0; Indels 100; Gaps 1;

Qy 1 MGKSDFLTTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGKSDFLTTPKAIANRIKSKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60

Qy 61 PQQFMDYFSEEFNDLFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDFTK 120
Db 61 PQQFMDYFSEEFNDLFLELLRRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDFTK 120

Qy 121 WLGRGLCKVDETPKGWYIQVIDRDPETIRROLEKLEKKKQDLDDDEKTAFFIEEQVRG 180
Db 121 WLGRGLCKVDETPKGWYIQVIDRDPETIRROLEKLEKKKQDLDDDEKTAFFIEEQVRG 180

Qy 181 LEGKEQVPTFTELSENDEKVTFNLSKGACSSSGATSSKSTLGPALKTIGSSASVK 240
Db 181 LEGKEQVPTFTELSENDEKVTFNLSKGACSSSGATSSKSTLGPALKTIGSSASVK 240

Qy 241 RKSSQSSQSKKKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVKIIITKLGKGYHK 300
Db 241 RKSSQSSQSKKKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVKIIITKLGKGYHK 200

Qy 301 KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAPGRILVLNGVGRNGEGTLESI 360
Db 301 KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAPGRILVLNGVGRNGEGTLESI 260

Qy 361 NEKTFSAITIVETGPKGRRVEGIOYEDISKLA 393
Db 261 NEKTFSAITIVETGPKGRRVEGIOYEDISKLA 293

```
RESULT 5
ID AAY23617
AC AAY23617 standard; protein; 291 AA.
XX
XX AAY23617;
DT 07-SEP-1999 (first entry)
XX
XX Mouse truncated kin17 protein.
XX
XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
KW protein interaction; curved DNA; HIV replication; HIV integration;
KW repair enzyme.
XX
XX Mus sp.
OS
PN PR2772046-A1.
XX
XX 11-JUN-1999.
PD
XX 09-DEC-1997; 97FR-00015536.
PF
XX 09-DEC-1997; 97FR-00015536.
PR
XX (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
PA
XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
PI WPI; 1999-359999/31.
XX
XX New DNA coding for human kin17 protein - useful for controlling cell
PT proliferation or fertility.
PT
XX Claim 14; Page 33-34; 69pp; French.
XX
XX The present sequence represents a mouse kin17 protein with amino acids
CC 129-228 deleted. The mammalian kin17 protein is useful for preparing a
CC medicament for controlling cell proliferation or for controlling
CC fertility. The medicaments can also be used to treat hyperproliferative
CC diseases. Fragments between amino acids 55 and 235 (preferably between
CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
CC regulating the interaction between proteins and curved DNA. The fragment
CC can be used to block replication of HIV or its integration into the human
CC genome or to target repair enzymes to curved DNA sites. Expression
CC vectors for kin17 can be used for controlling cell proliferation
XX
XX Sequence 291 AA;
SQ
Query Match 67.0%; Score 1365; DB 2; Length 291;
Best Local Similarity 69.2%; Pred. No. 4.6e-110;
Matches 272; Conservative 9; Mismatches 10; Indels 102; Gaps 2;
QY 1 MGKSDFLTPKAIANRIKSKGLQKLRWYCMQCQKQCRDENGFKCHCMSESHORQLLAS 60
DB 1 MGKSDFLSPKAIANRIKSKGLQKLRWYCMQCQKQCRDENGFKCHCMSESHORQLLAS 60
QY 61 PQOFMDYFSEEFNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHNNATQWETLDTFK 120
DB 61 PQOFMDYFSEEFNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHNNATQWETLDTFK 120
QY 121 WLGREGLCKVDYTPKGYIQYIDRDPETTRRQLELEKKKKQDLDDEEKTAKFIEEQVRG 180
DB 121 WLGREGLC----- 128
QY 181 LEGKEQVPTFTLSRENDEEKTVPNLSKGCSSSGATSSKSTLGPLSALKTIGSSASVK 240
DB 129 -----ALKLIGSAASGK 140
QY 241 RKSSQSSQSKKKKKKSGALDEIMBIEBKKTARTDYWLQPEIIVKIITTKLGEKYHK 300
DB 141 RKSSQSSAQ--PAKKKSGALDEIMELEBKKTARTDQWLQGVIVVVIITTKLGEKYHK 198
```

```
QY 301 KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAGKRILVNGGYRGNEGTFLESI 360.
DB 199 KKGVVKEVIDRYTAVVKMTDSGRDLKDQTHLETVIPAGKRVLVNGGYRGNEGTFLESI 258
QY 361 NEKTSATIVTETGPKGRRVEGQYEDISKLA 393
DB 259 NEKAFSATIVTETGPKGRRVEGQYEDISKLA 291
RESULT 6
ADP22452
ID ADP22452 standard; protein; 390 AA.
XX
XX ADP22452;
XX
XX 12-AUG-2004 (first entry)
XX
XX Sea-squirt (Ciona intestinalis) zinc finger protein #13.
DE
XX sea-squirt; zinc finger protein; gene detection; drug development;
KW zinc finger protein-associated disease.
XX
XX Ciona intestinalis.
OS
XX JP2004057126-A.
PN
XX 26-FEB-2004.
PD
XX 31-JUL-2002; 2002JP-00222484.
PF
XX 31-JUL-2002; 2002JP-00222484.
PR
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA
XX WPI; 2004-208711/20.
XX
XX N-PSDB; ADP22451.
DR
XX Novel gene encoding zinc finger protein, useful as probe in gene
PT detecting instruments and in development of drug for treating zinc finger
PT protein associated diseases.
XX
XX Claim 1; SEQ ID NO 26; 972pp; Japanese.
PS
XX The invention comprises the amino acid and coding sequences of sea-squirt
CC (Ciona intestinalis) zinc finger proteins. The DNA and protein sequences
CC of the invention are useful in a gene detecting instrument. The DNA and
CC protein sequences of the invention are useful in the development of drugs
CC for the treatment of zinc finger protein-associated diseases. The present
CC amino acid sequence represents a sea-squirt zinc finger protein of the
CC invention.
XX
XX Sequence 390 AA;
SQ
Query Match 61.0%; Score 1243; DB 8; Length 390;
Best Local Similarity 62.5%; Pred. No. 2.9e-99;
Matches 247; Conservative 48; Mismatches 86; Indels 14; Gaps 6;
QY 1 MGKS--DFLTPKAIANRIKSKGLQKLRWYCMQCQKQCRDENGFKCHCMSESHORQLLAS 58
DB 1 MGKEKPGFLITAKIAGNRIKSKGLQKLRWYCMQCQKQCRDENGFKCHCMSESHORQLIAG 60
QY 59 ENPOQFMDYFSEEFNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHNNATQWETLTF 118
DB 61 ENPGAFHSFFSSFPFQMLKTRFGTKRVHNNIVYNEYISHREHIHNNATKWTLTAF 120
QY 119 TKWLREGLCVKDYTPKGYIQYIDRDPETTRRQLELEKKKKQDLDDEEKTAKFIEEQVR 178
DB 121 TKWLREGHCKVDQTEKGFQYIYDRDPLVLARQLELKQKQREDDDERQAKAIEEMVK 180
QY 179 RGLG-GKEQEVPTFTLSRENDEEKTVPNLSKGCSSSGATSSKSTLGPLSALKTIGSSA 237
DB 181 RQDQTAKSDSGKFTLVLR-NDEEKIAINLQ---MNSEPKAATSSQLAPRV-----SLS 230
```

```
QY 238 SVRKSSQSTOSKEKKKKSALDEIMEIEE-EKKRTARTDYWLOPEIIVKIITKIGE 296
Db 231 SLVGKKQPSSEKFSKRSKKSALDEIAAEBGRKKQKLHTEHLAKNIVVKIITKIGD 290
QY 297 KYHKKAIVKEVIDKYTAVVMIDSGDKLKDQTHLETVIPAPGRILVNGGYRNEG 356
Db 291 KYFKKXAVKEVDKFTAI VKMIDSGDVQKFDQTHLETVIPSGTKTVLI VNGGYRDCVAT 350
QY 357 LESINEKTPSATIVETGPKLGRVEGIQYEDISK 391
Db 351 LEKIDTSKFSATLKITEGPLNGRIVEDIKYEDFSK 385

RESULT 7
ABB71416
ID ABB71416 standard; protein; 390 AA.
AC ABB71416;
XX
XX
DT 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 41040.
DE
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL15519.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 41040; 2lpp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX sequences (AB101840-AB116175) and the encoded proteins (ABBS7737-
XX ABBS72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 390 AA;
XX
XX Query Match 52.0%; Score 1059.5; DB 4; Length 390;
XX Best Local Similarity 50.7%; Pred. No. 2.8e-83;
XX Matches 207; Conservative 73; Mismatches 91; Indels 37; Gaps 7;
XX
XX 1 MKGSDFLTPKAIANRIKSGLOKRWYQCMQCKQCRDENGFKCHMSHSHORQLLAS 60
XX
XX 1 MGRAEVGTPKYLANKRKSKGLQKRWYQCMQCKQCRDENGFKCHTMSHSHORQLL 60
XX
XX 61 PQQFMDYSESEFRNDLFLLRRRTGKRVHNNIVYNEVISHREHIMNATOWETLTD 120
XX
```

```
Db 61 PGKFLHSFSESGVMYELLRRRFGTKRTSANKIYQEIYIAHKEHIMNATRWLTLSDYVK 120
QY 121 WLGRGLCKVDTPKGYIOYIDRPETIRROLELEKKKKQDLDDREKTAKTFIEQVR-- 178
Db 121 WLGRGQVADTEKGFVYIDRPFAMERQAKADRKEMKEDDERVADFIEQIKNV 180
QY 179 RGLGKQEQE-VPTFTLSRENDEE-KVTNLSKAGACSSSGATSSKSTGLPSALKTIGSS 236
Db 181 KAKDGEDEGQKFTLKEENEPLKDIRLEK-----KFQPDTVLGSAL----- 226
QY 237 ASVKRKSSQSTOSKEKKKKK-----SALDEIMEIEEKK-RTARTDYWLOPE 284
Db 227 -----AKRPAPAEAEKVKPKSVAGDSQTSRSLDEIIKOEESKKEANRKYDLHKG 279
QY 285 IIVKIITKLGKGYHKKKAIKVEVIDKYTAVVMIDSGDKLKDQTHLETVIPAPGRIL 344
Db 280 IIVKFKSKMGKFPKQKAVLDVDRYQKIKFLETGKLVQDAHLETVIPALDKPVM 339
QY 345 VLNGGYRNEGTLSEINEKTPSATIVETGPKLGRVEGIQYEDISK 392
Db 340 VVNGAYRGSEALLRKLDERRYSVSVELHGLPLKGRIVDNVQVEDISK 387

RESULT 8
AAG42576
ID AAG42576 standard; protein; 411 AA.
XX
XX AAG42576;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 53114.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EF1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX
XX 05-MAR-1999; 99US-0123180P.
XX
XX 09-MAR-1999; 99US-0123548P.
XX
XX 23-MAR-1999; 99US-0125788P.
XX
XX 26-MAR-1999; 99US-0126264P.
XX
XX 29-MAR-1999; 99US-0126785P.
XX
XX 01-APR-1999; 99US-0127462P.
XX
XX 06-APR-1999; 99US-0128234P.
XX
XX 08-APR-1999; 99US-0128714P.
XX
XX 16-APR-1999; 99US-0129845P.
XX
XX 19-APR-1999; 99US-0130077P.
XX
XX 21-APR-1999; 99US-0130449P.
XX
XX 23-APR-1999; 99US-0130510P.
XX
XX 28-APR-1999; 99US-0130891P.
XX
XX 30-APR-1999; 99US-0131449P.
XX
XX 30-APR-1999; 99US-0132048P.
XX
XX 30-APR-1999; 99US-0132407P.
XX
XX 04-MAY-1999; 99US-0132484P.
XX
XX 05-MAY-1999; 99US-0132485P.
XX
XX 06-MAY-1999; 99US-0132486P.
XX
XX 07-MAY-1999; 99US-0132487P.
XX
XX 11-MAY-1999; 99US-0132863P.
XX
XX 14-MAY-1999; 99US-0134256P.
XX
XX 14-MAY-1999; 99US-0134218P.
XX
XX 14-MAY-1999; 99US-0134219P.
XX
XX 14-MAY-1999; 99US-0134221P.
XX
XX 14-MAY-1999; 99US-0134370P.
XX
XX 18-MAY-1999; 99US-0134768P.
```

PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142520P.
PR 12-JUL-1999; 99US-0142377P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.

PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158023P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159323P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.

```

PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 48.08; Score 978; DB 3; Length 411;
Best Local Similarity 48.68; Pred. No. 3.8e-76;
Matches 201; Conservative 66; Mismatches 123; Indels 24; Gaps 6;

Qy 1 MGKSDFTLPKAIANRIKSGIQLRWYQCMQCRDENGFKCHMSHQRQLLASEN 60
Db 1 MGKNDFTLPKAIANRIKAKGLQKLRWYQCMQCRDENGFKCHMSHQRQMVFGQN 60

Qy 61 PQQMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHHMNATOWETLDTFK 120
Db 61 PTRVVDGYSEFEQTFDLRRSHRFSRIAATVVYNEYINDRHVHNMSTEWATLTFFIK 120

Qy 121 WLGRGLCKVDETPKGWYIOYIDRDPETIRROLELEKKKKQDLDDDEKTAFFIEQVRRG 180
Db 121 HLKGTGCKVEETPKGWFYIYDRDSEAVKRLKRKIKSDMADDERQRMIAQIERA 180

Qy 181 LE-----GKEQEVPTFELSGRENDEE-----KVTFNLSKGACS--SSGATSSK 221
Db 181 AEKLNCGGGEGETSGNDEVDDGDERKXDEDLRLKSGVKGFALGGVQVATGKERGE 240

Qy 222 SSTLPSALKITIGSSASVKRESQSSTQSKBKKKKXSALDEIMEIEEEKK-RTARTDYW 280
Db 241 SSKLLFGDEENDKVERGEKRRSGDS---GRSEKERRSALDELMEKEBKKKERNMNRKDYW 297

Qy 281 LQPIIIVKIITKLGEEK-YHKKAIVKEVIDKYTAVVVMIDSGDKLKDQTHLETVPAP 339
Db 298 LFEGIIIVKMSKALAEGYKQKGVVKKVVDNYVGEIKMLDSKHLVLRVQKLETLVLPQI 357

Qy 340 GKRLVLNGGYRGNEGTLNESINEKTFSATIVETGPKLGRVVEIGIQVEDISKLA 393
Db 358 GGMVKIVNGAYRGSNARLLGVDTKFCVKQIEKGVVDGRVKSIEYEDICKLA 411

RESULT 9
AAB03063
ID AAB03063 standard; protein; 424 AA.
XX AC AAB03063;
XX DT 27-SEP-2000 (first entry)
XX DE Maize KIN17 orthologue, ZmKINH-1.
XX ZmKINH-1; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
KW nuclear localisation; nonhomologous recombination;
KW illegitimate recombination; double stranded DNA binding; curved DNA;
KW homologous gene targeting; transgenic plant.
XX OS Zea mays.
XX PN WO200024900-A1.
XX PD 04-MAY-2000.
XX PP 06-OCT-1999; 99WO-US023280.
XX PR 27-OCT-1998; 98US-0105802P.
XX PA (PION-) PIONEER HI-BRED INT INC.
XX PI Mahajan PB;
XX WPI; 2000-350741/30.
XX DR N-PSDB; AAS52589.
XX Nucleic acids encoding maize KIN17 orthologue proteins useful for

```

preventing illegitimate recombination in cells.

Claim 11; Page 64-66; 84pp; English.

This sequence represents the maize KIN17 orthologue ZmKINH-1. The invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic acids encoding them (AAS52589-A52591), and expression vectors, transgenic plants and plant seeds comprising nucleotides encoding maize KIN17 orthologues. KIN17 has, until now, been found only in animal (avian, rodent and human) cells, this invention being the first report describing the presence of KIN17 in plants. Murine KIN17 was found to have a significant homology to Escherichia coli RecA protein, and contains a zinc finger motif and a nuclear localisation signal. KIN17 binds double-stranded DNA, preferentially binding to curved DNA, and forms intranuclear foci on overexpression in mammalian cells. It is also induced on exposure to gamma or ultraviolet radiation. These findings indicate that KIN17 plays a role in non-homologous (illegitimate) recombination, which occurs at higher rates among higher eukaryotes, particularly plants. Illegitimate recombination in plants is a major impediment to the generation of transgenic crops such as maize. Maize KIN17 orthologue nucleic acid sequences may be used to generate transgenic plants. The transgenic plants generated can be monocots or dicots and are particularly maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense sequences may be used to reduce KIN17 levels in embryogenic callus or embryo cells, thereby reducing the amount of non-homologous recombination and enhancing homologous gene targeting

Sequence 424 AA;

Query Match 45.9%; Score 934; DB 3; Length 424;
Best Local Similarity 45.3%; Pred. No. 2.7e-72;
Matches 193; Conservative 75; Mismatches 122; Indels 36; Gaps 7;

```

Qy 1 MGKSDFTLPKAIANRIKSGIQLRWYQCMQCRDENGFKCHMSHQRQLLASEN 60
Db 1 MGKHEFTLPKAIANRIKAKGLQKLRWYQCMQCRDENGFKCHMSHQRQMVFGMA 60

Qy 61 PQQMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHHMNATOWETLDTFK 120
Db 61 PDRVVEGFSEEFLESLIRRAHRSRVAAVVYNEYIADRHVHNMSTRWATLTFFVK 120

Qy 121 WLGRGLCKVDETPKGWYIOYIDRDPETIRROLELEKKKKQDLDDDEKTAFFIEQVRR- 179
Db 121 FLGREGYCKVEDTPKGWFYIYDRDSEAVKRLKRKIKSDMADDERQRMIAQIERA 180

Qy 180 -----GLEGEQEVPTFTELSENDEEKVTFNLSKGACSSG--ATSSKSTLGP 227
Db 181 HKSLAKPNGGAAEGEPESGEYSGDDDEQSDDEADKATGKIALQKAVPGP 240

Qy 228 SALKTIGSSASVKRK-----ESSQSSSQSKKKK-----KKSALDEIMEIEE 269
Db 241 ---KVPFDDPKPKMKFGFDEEDSGTDRQSKNELTKMGKDVKAAEAKRSALDLMKEEE 297

Qy 270 E-KKRTARTDYWLOPEIIVKIITKLGEEK-YHKKAIVKEVIDKYTAVVVMIDSGDKLKL 327
Db 298 KAKERSNRKDYWLCPGIVVVKMSKLAEGYKQKGVVKKVVDNYVGEIKMLDSKHLVLRV 357

Qy 328 DQTHLETVPAPGKRILVNLNGYRGNEGTLNESINEKTFSATIVETGPKLGRVVEIGIQYE 387
Db 358 DQDELETVPQIIGGLVRVNGAYRGSNARLLGVDTKFCVKQIEKGVVDGRVKSIEYEDICKLA 417

Qy 388 DISKLA 393
Db 418 DICKIS 423

```

RESULT 10
ADX96318
ID ADX96318 standard; protein; 437 AA.

XX AC ADX96318;
XX

XX New recombinant DNA constructs useful in the field of biochemistry and
PT genetics, and in particular for producing transgenic plants with improved
PT biological characteristics.
XX
XX
PS Claim 2; SEQ ID NO 8499; 14pp; English.
XX
CC The invention relates a recombinant DNA construct comprising a
CC polynucleotide having any of 5544 nucleotide sequences (CDNAs SEQ ID NO:
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences
CC (SEQ ID NO: 5545-11088). The CDNAs and proteins are from corn, soybean,
CC Arabidopsis, wheat and rape but the specification does not indicate which
CC sequences is derived from which organism. Also included is a method of
CC producing a plant having an improved property, comprising transforming a
CC plant with a recombinant DNA construct comprising a promoter region
CC functional in a plant cell operably joined to a polynucleotide encoding a
CC polypeptide associated with the property, and growing the transformed
CC plant. The property is selected from improving plant cold tolerance, for
CC manipulating growth rate in plant cells by modification of the cell cycle
CC pathway, for improving plant drought tolerance, for providing increased
CC resistance to plant disease, for galactomannan production, for production
CC of plant growth regulators, for improving plant heat tolerance, for
CC improving plant tolerance to herbicides, for increasing the rate of
CC homologous recombination in plants, for lignin production, for improving
CC plant tolerance to extreme osmotic conditions, for improving plant
CC tolerance to pathogens or pests, for yield improvement by modification of
CC photosynthesis, for modifying seed oil yield and/or content, for
CC modifying seed protein yield and/or content, for yield improvement by
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake
CC and for yield improvement by providing improved plant growth and
CC development under at least one stress condition. The polynucleotide may
CC also encode a plant transcription factor. The methods and compositions of
CC the present invention are useful in the field of biochemistry and
CC genetics, in particular for producing transgenic plants with improved
CC biological characteristics such as increased yield, improved nitrogen
CC flow, increasing plant tolerance to cold or heat, improving plant
CC tolerance to extreme osmotic and drought conditions, and improving plant
CC tolerance to plant pests or pathogens. They can also be used in physical
CC arrays of molecules, plant breeding markers, computer-based storage and
CC analysis systems. The present sequence is one of the 5544 plant protein
CC sequences of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.
XX
XX Sequence 423 AA;

Query Match 44.6%; Score 909.5; DB 8; Length 423;
Best Local Similarity 44.6%; Pred. No. 3.6e-70;
Matches 189; Conservative 76; Mismatches 126; Indels 33; Gaps 8;

QY 1 MGKSDFLTPKAIANRIKSGKQLRWYCOMQKQCRDENGFKCHCMSSHQRQLLASEN 60
DB 1 MGKHEFLTPKAIANRIKAKGKQLRWYCOMQKQCRDENGFKCHCMSSHQRQVFGMA 60
QY 61 PQQFMDYSEFRNDLFLRRRFTKRVHNNIVYNEIYSHREHHMATOWEITLDTFK 120
DB 61 PDRVVEGSEIFLESFLSLIRRAHRSVAATVVVYNEIADRHVHNNSTRMATLTFEVK 120
QY 121 WLGRGLCKVDTPKGYIQYIDRDPETIRQLELEKKKKODDDEEKTAKFIEQVR- 179
DB 121 LLGREGYCKVETPKGWFTYIDRSEQAVKDLRKRIKSDMAEDERQERMIARQIERA 180
QY 180 -----GLEGEQEVPTFTLSRENDEBEKVTFNLSKGCSSSG--ATSSKSSITLG 226
DB 181 HKS LAKPNDGDAEG-EPESGEEYVSGDDGEELEDVSEKANKATGKIALQATVRG 239
QY 227 PSALTKTIGSSASVKKSSQSSTQSEKK-----KKK-----SALDEIMEIER-E 270
DB 240 PK-VNPFEDKPKVRFDEGLTGCEBKDELAKKKGDKVDKAAARRRSGALDELMEEMA 298
QY 271 KRTARTDYWLOPEIIVKTIITKLGK-YHKKAIVKEVIDKYTAVVKMIDSGDKLKDQ 329

Db 299 KERSNRKDYMLCPGIVVVKVMSKSLAEKGYKQKGVVRKVMVKVGEIEMLESKHVLRVDQ 358
QY 330 THLETVIPAGKRILLNGSYRGNEGTLESINEKTSATIVISTGPKGRRVREGIOVEDI 389
Db 359 DELETVIPQIGGLVRLINGAYRGSNARLLSVDTKECAKQVVEKGLYDGKVLRAVKYDI 418
QY 390 SKLA 393
Db 419 CKIS 422
RESULT 15
AAG42577
ID AAG42577 standard; protein; 382 AA.
XX AAG42577;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 53115.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 25-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 08-APR-1999; 99US-0128234P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 28-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 07-MAY-1999; 99US-0132863P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 18-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134768P.
XX 19-MAY-1999; 99US-0134941P.
XX 20-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135353P.
XX 25-MAY-1999; 99US-0135629P.
XX 27-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
XX 28-MAY-1999; 99US-0136782P.
XX 01-JUN-1999; 99US-0137222P.
XX 03-JUN-1999; 99US-0137528P.
XX 04-JUN-1999; 99US-0137502P.
XX 07-JUN-1999; 99US-0137724P.
XX 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 28-JUN-1999; 99US-0140891P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 08-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143342P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149428P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157533P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 40.7%; Score 829; DB 3; Length 382;
Best Local Similarity 45.2%; Pred. No. 3.2e-63;
Matches 174; Conservative 64; Mismatches 123; Indels 24; Gaps 6;
YQ 30 MCQKQCRDENGFKCHMSHQRQLLASENPQQFMDYFSEBFNDFLLELRRTGTRV 89

```
Db      1 MCQKQCRDENGFKCHCSESHQRMQVFGQNPTRVVDGYSEEPQTFDLMRRSHRESRI 60
Qy      90 HNNIVYNEYISHREHHHNNATOWETLTDFTKWLGREGLCCKVDETPKGYIOYIDRDPETI 149
Db      61 AATVVYNEYINDRRHHVHMNSTEWATLTETFKHLGKTGCKVEETPKGWFITYIDROSETL 120
Qy      150 RROLELEKKKKQDLDDDEEKTAKFTEEQVVRGLE-----GKEQEVPTFTETLSRENDEE--- 201
Db      121 FKRLKNKRVKSDJAEBEKKQREIQRIERAAEKLNGGGGEGETSGNDEVVDGDDDERKK 180
Qy      202 -----KVTFNLSKGACS-SSGATSSKSSSTLGPALKTTIGSSASVKKRESSQSSSTQ 250
Db      181 DEDLRLKSGVKVGFALGGGVQVATKGERGESSKLLFGDEENDKVERGEKKRSGDS--- 237
Qy      251 SKEKKKKKSADEIMEIEBEKK-RTARTDYLWQPEIIVKIITKKLGEBK-YHKKAIVKEV 308
Db      238 GRSEKERRSALDELKMBEKKERMNRKDYWLFEIIVKVMKALAEKGYKQKGVVYKV 297
Qy      309 IDKYTAVVVMIDSGDKLKDQTHLETVI PAPKRIILVNGGYRGNEGTLESINEKTFSAT 368
Db      298 IDNTVGEIKMLDPSKHVLKVDQKELETVLQITGGMVKIIVNGAYRGSNARLLIGVDTEKFAK 357
Qy      369 IVIETGPLKRRRVEGIOVEDISKLA 393
Db      358 VOIEKGVYDGRVIXSIEYEDICKLA 382
```

Search completed: November 23, 2005, 16:28:22
Job time : 132.833 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:19:43 ; Search time 153.39 Seconds
(without alignments)
1807.628 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 2037
Sequence: 1 MGKSDFLTKAIANRIKSG.....GPLKGRVGEIQVEDISKLA 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2037	100.0	393	2	O60870 HUMAN
2	1895	93.0	391	2	Q8K339 mus musculus
3	1702	83.6	389	2	Q6GL97 XENTR
4	1653	81.1	387	2	Q68F56 xenopus lae
5	1568	77.0	383	2	Q5RH25 BRARE
6	1114	54.7	387	2	Q7PGA1 ANOGA
7	1059.5	52.0	390	2	Q9VPH4 DROME
8	1057.5	51.9	390	2	Q8SXR2 DROME
9	1044	51.3	195	2	Q9CV58 MOUSE
10	978	48.0	411	2	Q9ZVU5 ARATH
11	952.5	46.8	404	2	Q9XWF2 CAEEL
12	936.5	46.0	269	2	Q4TDV9 TETNG
13	918	45.1	430	2	Q75LU5 ORYSA
14	849.5	41.7	427	2	Q4S582 TETNG
15	845	41.5	402	2	Q9AY87 ORISA
16	833	40.9	445	2	Q55D16 DICDI
17	830	40.7	397	2	Q60L39 CAEBR
18	694.5	34.1	244	2	Q76926 DROME
19	659	32.4	378	2	Q4N9K7 THEPA
20	655	32.2	441	2	Q4XDV6 PLACH
21	655	32.2	445	2	Q4Z7D7 PLABE
22	653	32.1	178	2	Q6X190 DROYA
23	652	32.0	412	2	Q4UG63 THEAN
24	646	31.7	442	2	Q8IKG2 PLAF7
25	640	31.4	457	2	Q7RD49 PLAYO
26	593.5	29.1	334	2	Q4WFG0 ASPFU
27	591	29.0	337	2	Q415W6 GIBZE
28	586.5	28.8	304	2	Q9Y7X9 SCHPO
29	586.5	28.8	335	2	Q5AQSO EMENI
30	541	26.6	324	2	Q55M94 CRYNE
31	541	26.6	324	2	Q5K8H7 CRYNE

32	535.5	26.3	328	2	Q525W2 MAGGR
33	522	25.6	358	2	Q4PG11 USTWA
34	506.5	24.9	256	2	Q6C2K1 YARLI
35	449.5	22.1	276	2	Q7S742 NEUCR
36	437.5	21.5	263	2	Q5CFM3 CRYHO
37	436	21.4	265	2	Q5CYD1 CRYPV
38	405.5	19.9	271	2	Q6BQO0 DEBHA
39	395.5	19.4	238	2	Q51F62 ENTHI
40	379.5	18.6	236	2	Q759P8 ASHGO
41	370	18.2	239	2	Q6CXJ7 KLULA
42	369	18.1	239	2	Q5A673 CANAL
43	365	17.9	121	2	Q4XND8 PLACH
44	352	17.3	241	2	Q6FMS8 CANGA
45	314.5	15.4	88	2	Q4TDW0 TETNG

ALIGNMENTS

RESULT 1

O60870 HUMAN
ID O60870_HUMAN PRELIMINARY; PRT; 393 AA.
AC O60870;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-JAN-1999 (Tremblrel. 09, Last sequence update)
DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)
DE Kin17 protein (HsKin17 protein) (KIN, antigenic determinant of reca
protein homolog) (Mouse).
GN Name:KIN; Synonyms:Kin17; ORFNames:RP11-264C14.1-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Testis;
RX MEDLINE=20419742; PubMed=10964102; DOI=10.1093/carcin/21.9.1701;
RA Kannouche P., Mauffrey P., Pinon-Lataillade G., Mattei M.G.,
Sarasin A., Daya-Grosjean L., Angulo J.F.;
RT "Molecular cloning and characterization of the human KIN17 cDNA
encoding a component of the UVC response that is conserved among
metazoans.";
RL Carcinogenesis 21:1701-1710(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Colling F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S., Krzywinski M.I., Skalska U., Small D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Pancreas;
RX Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBSJ databases.

```

[4]
RP NUCLEOTIDE SEQUENCE.
RA Lovell J.;
RA Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ005273; CAA06462.1; -; mRNA.
DR EMBL: BC017309; AAH17309.1; -; mRNA.
DR EMBL: AL158044; CAI12959.1; -; Genomic DNA.
DR Ensemble; ENSG00000151657; Homo sapiens.
DR HGNC: HGNC:6327; KIN.
DR GO: GO:0005634; C:nucleus; TAS.
DR GO: GO:003677; F:DNA binding; TAS.
DR InterPro: IPR005824; KOW.
DR InterPro: IPR007087; Znf_C2H2.
DR Pfam: PF00467; KOW; 1.
DR SMART; SM00355; ZNF_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ SEQUENCE 393 AA; 45374 MW; 515A89B4C8A4C007 CRC64;

Query Match 100.0%; Score 2037; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 2.2e-104;
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKSDFLTPKAIANRIKSKGLQRLWYCMQCKQCRDENGFKCHCWSHQRQLLASEN 60
DB 1 MGKSDFLTPKAIANRIKSKGLQRLWYCMQCKQCRDENGFKCHCWSHQRQLLASEN 60
QY 61 PQQFMDYFSEPRNDFLELLRRRFGTKRVHNNIVVYNEYISHREHIHMNATQWETLDTFK 120
DB 61 PQQFMDYFSEPRNDFLELLRRRFGTKRVHNNIVVYNEYISHREHIHMNATQWETLDTFK 120
QY 121 WLGRGLCKVDTPKGYIYIDRDPETIRROLELEKKKKQDLDDBEKTAFTIEEVRRG 180
DB 121 WLGRGLCKVDTPKGYIYIDRDPETIRROLELEKKKKQDLDDBEKTAFTIEEVRRG 180
QY 181 LEGKEQVPTFTELSRENDEEKTVPNLKSGACSSGATSSKSTGLPSSALKTIGSSASVK 240
DB 181 LEGKEQVPTFTELSRENDEEKTVPNLKSGACSSGATSSKSTGLPSSALKTIGSSASVK 240
QY 241 RKSSQSSSTQSKKKKKKSSALDEIMEIEBEKKRTARTDYWLOPEIIVKIITKLGEKYHK 300
DB 241 RKSSQSSSTQSKKKKKKSSALDEIMEIEBEKKRTARTDYWLOPEIIVKIITKLGEKYHK 300
QY 301 KKAIVKEVDKYTAVVKMIDSGDKLKDQTHLETVPAPGKRIIVLNGGYRGNEGTLSEI 360
DB 301 KKAIVKEVDKYTAVVKMIDSGDKLKDQTHLETVPAPGKRIIVLNGGYRGNEGTLSEI 360
QY 361 NEKTSATVIETGPKLGRVEGQYEDISKLA 393
DB 361 NEKTSATVIETGPKLGRVEGQYEDISKLA 393

RESULT 2
O8K339 MOUSE
ID Q8K339_MOUSE PRELIMINARY; PRT; 391 AA.
AC Q8K339;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Antigenic determinant of rec-A protein.
GN Names:kin;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

```



```
DR GO: 0008270; F: zinc ion binding; IEA.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00467; KOW; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR SEQUENCE 387 AA; 44491 MW; 4CCB56B047F3779 CRC64;

Query Match      81.1%; Score 1653; DB 2; Length 387;
Best Local Similarity 78.9%; Pred. No. 3.1e-83;
Matches 310; Conservative 45; Mismatches 32; Indels 6; Gaps 4;

QY 1 MGKSDFLTPKAIANRIKSKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQROLILASEN 60
DB 1 MGKADFLSPKAIANRIKSKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQROLILASEN 60
QY 61 PQQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIMHNATQWETLTDFTK 120
DB 61 PQQIMDSFSEFTFLELLRRRFGTKRVHNNIVYNEYIGHREHVMHNATQWETLTDFTK 120
QY 121 WLGRGLCKVDPTPKGWYIQYIDRDPETIRRQLELEKKKKQDLDDDEKTAKEFIEEQVRRG 180
DB 121 WLGRGLCKVDPTPKGWYIQYIDRDPETIRRQLELEKKKKQDLDDDEKTAKEFIEEQVRRG 180
QY 181 LEGKE-QEYPTFTELSRENDEEKVTNLSKGACSSSGATSSKSTLGPSSALKTIGSSASV 239
DB 181 REGREPTPEVPTELQKSEDEKVAFLNLEIAA---AGPSKPSTTLGPSALKT---TASV 233
QY 240 KKESQSSQSTQSEKKEKKKKSALDEIMEIEEKKRTARTDYWLOPEIIIVKIITKLGEKYH 299
DB 240 KKESQSSQSTQSEKKEKKKKSALDEI---:---:---:---:---:---:---:---: 289
QY 300 KKAIVKEIDKYTAVVMKIDSGDKLKDQTHLETVPAPGRKRLVNLGVYRGNEGTLES 359
DB 300 KKAIVKEIDKYTAVVMKIDSGDKLKDQTHLETVPAPGRKRLVNLGVYRGNEGTLES 359
QY 360 INEKTFSATIVETGLKGRVREGIQYEDISKLA 393
DB 360 INEKTFSATIVETGLKGRVREGIQYEDISKLA 393
DB 350 INEHFSAITLUDSGRMKGKTVEGIAVEDFSKLA 383

RESULT 6
Q7FGAI ANOQA
ID Q7FGAI ANOQA PRELIMINARY; PRT; 387 AA.
AC Q7FGAI.
DT 01-MAR-2004 (TremBLrel. 26, Created)
DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE ENSANGP0000023646 (Fragment).
GN ORFNames=ENSANGG0000020592;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG "Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01008849; EAA45011.2; -; Genomic_DNA.
DR InterPro; IPR007087; Znf_C2H2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 387 AA; 44275 MW; F1B6FE04131DD85D CRC64;

Query Match      54.7%; Score 1114; DB 2; Length 387;
Best Local Similarity 54.8%; Pred. No. 1.5e-53;
Matches 216; Conservative 67; Mismatches 99; Indels 12; Gaps 3;

QY 2 GKSDFLTPKAIANRIKSKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQROLILASEN 61
DB 1 GKAEVGTPKYLANKAKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQROLILASEN 60
QY 62 QQFMDYFSEFRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIMHNATQWETLTDFTK 121
DB 61 GRFIDGFSFELTQVILRRQFGTKRVAANKVYQEYIADRHLLHMNATKMHSLSDFKY 120
QY 122 LGREGGLCKVDPTPKGWYIQYIDRDPETIRRQLELEKKKKQDLDDDEKTAKEFIEEQVRRGL 181

GO: 0008270; F: zinc ion binding; IEA.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00467; KOW; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR SEQUENCE 387 AA; 44491 MW; 4CCB56B047F3779 CRC64;

Query Match      77.0%; Score 1568; DB 2; Length 387;
Best Local Similarity 75.9%; Pred. No. 1.5e-78;
Matches 299; Conservative 41; Mismatches 42; Indels 12; Gaps 5;

QY 1 MGKSDFLTPKAIANRIKSKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQROLILASEN 60
DB 1 MGKADFLTPKAIANRIKAKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHQROLILASEN 60
```



```
Db 121 LGRNGCHVADTDKGFITIDRDPEFLANQKQKQKDDASELAEFIEQVRGK 180
Qy 182 EGKEQEVPTTELSRENDEKVTNLSKGACSSGATSKSTLGPAL---KTIGSSASV 239
Db 181 TEBEPCSTGVSSELKRENEEDTIKIELKLG-----SKQOOSTPSAVISKRPFALDD 231
Qy 240 KRKESSOSTQSKKKKKKALDEIMEIEBEKK-RTARTDYWLQPEITIVKIITKLGKEY 298
Db 232 GKXKKTIAATSTNGETKLSALDELIOEBEQKKEKNRKDYWLAEGLVVKLRSLGKEY 291
Qy 299 HKKAIKVEIDKTYAVVMKIDSGDKLKLDQTHLETIVPAGKRILVINGGYRGNEGLE 358
Db 292 YKEGVVVEIEKRYAKIKLETGEKLVQDAHLETIVPAGKQILVINGGYRGCTAVLK 351
Qy 359 SINEKTSATIVTETPLKGRVREGIOYEDISKL 392
Db 352 AINTERYSVTETIASGPKLGRVSNVAYEDISKL 385

RESULT 7
Q9VPH4 DROME PRELIMINARY; PRT; 390 AA.
AC Q9VPH4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CS5649-PA.
GN Names=kin17; ORFNames=CG5649;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN NUCLEOTIDE SEQUENCE.
[1]
MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heinen T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laake P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
```

```
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskaas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RN Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirskaas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective.";
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Battencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RN Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacle J., Park S., Svirskaas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP NUCLEOTIDE SEQUENCE.
RG FlyBase;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
CC -!- INTERACTION:
CC QSVI16:CG1041; NbExp=1; IntAct=EBI-193999, EBI-175006;
CC QVPS5:Hap60B; NbExp=1; IntAct=EBI-193999, EBI-159512;
CC P1187:pAbp; NbExp=1; IntAct=EBI-193999, EBI-103658;
DR EMBL; AE003591; AAF51578.1; -; Genomic_DNA.
DR IntAct; Q9VPH4; -;
DR Ensembl; CG5649; Drosophila melanogaster.
DR FlyBase; FBgn0024887; CG5649.
DR FlyBase; FBgn0024887; kin17.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
DR Nuclear protein.
SQ SEQUENCE 390 AA; 45355 MW; 5F8231706C2D431F CRC64;
Query Match 52.0%; Score 1059.5; DB 2; Length 390;
Best Local Similarity 50.7%; Pred. No. 1.5e-50;
Matches 207; Conservative 73; Mismatches 91; Indels 37; Gaps 7;
Qy 1 MGKSDFTTPKAIAIRIKSKGLQKLRTWCQKQCRDENGFKCHMCESHQRQLLASEN 60
Db 1 MGRAEVGTGPKYLANRMSKGLQKLRTWCQKQCRDENGFKCHMCESHQRQLLAFADN 60
```

```
QY 61 PQOFMDYFSEERNDPFLRLRRFGTKRVHNNIVYNEYISHREHHIMNATQWETLTDFTK 120
DB 61 PGKFLHSFSEKSDGMYELLRRFRFGTKRTSANKIYQYIAHKEHHIMNATRWLTLSDYVK 120
QY 121 WLGRGLCKVDTPKGYIOYIDRDPETIRROLELEKXKQDLDEKTAKEIEBOVR-- 178
DB 121 WLGRTCQVIADETEKGFVTVYIDRSPEAMERQAKADKEMKEXKODDERWADFTIEQIKNA 180
QY 179 RGLGKGEQEB-VPTFTLSRENDEE-KVTFNLKSGACSSSGATSSKSTLGPALKITIGSS 236
DB 179 RGLGKGEQEB-VPTFTLSRENDEE-KVTFNLKSGACSSSGATSSKSTLGPALKITIGSS 236
QY 181 KAKDGEEDGQKFTLKRRENEPLKDIRLEK-----KFQPDVTLGKSAL----- 226
DB 181 KAKDGEEDGQKFTLKRRENEPLKDIRLEK-----KFQPDVTLGKSAL----- 226
QY 237 ASVKRKESSQSTQSKKKKK-----SALDEIMEIEEEKK-RTARTDYWLQPE 284
DB 237 ASVKRKESSQSTQSKKKKK-----SALDEIMEIEEEKK-RTARTDYWLQPE 284
QY 227 -----AKRPAPEAEKVKPKSVAGDSQTRSVDLDEIIKQESKKEKRAKRDYWLHKG 279
DB 227 -----AKRPAPEAEKVKPKSVAGDSQTRSVDLDEIIKQESKKEKRAKRDYWLHKG 279
QY 285 IIVKIITKLGEYHKYKKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVTPAPCKRIL 344
DB 285 IIVKIITKLGEYHKYKKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVTPAPCKRIL 344
QY 280 IIVKFTSKMGKFFKQKAVLVLDVDRYQKIKFLETGKLVQDQAHLETVPALDKPVM 339
DB 280 IIVKFTSKMGKFFKQKAVLVLDVDRYQKIKFLETGKLVQDQAHLETVPALDKPVM 339
QY 345 VLVGGYRGNEGTLSEINETSATIVETGPLKGRVVEGIQYEDISK 392
DB 345 VLVGGYRGNEGTLSEINETSATIVETGPLKGRVVEGIQYEDISK 392
QY 340 VVNGAYRGSEALLKLDERRYSVSVLEIHLGPKLGRIVDNVQYEDISK 387
DB 340 VVNGAYRGSEALLKLDERRYSVSVLEIHLGPKLGRIVDNVQYEDISK 387

RESULT 8
Q8SRX2 DROME
ID Q8SRX2 DROME PRELIMINARY; PRT; 390 AA.
AC Q8SRX2
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE Mus musculus adult male tongue cDNA, RIKEN full-length enriched
DE library, clone:2310020017 product:antigenic determinant of rec-A
DE protein, full insert sequence. (fragment).
DE Name=Kin;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Berkeley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dreesnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY084184; AAL89922.1; -; mRNA.
DR FlyBase; FBgn0024887; CG5649.
DR FlyBase; FBgn0024887; Kin17.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003676; P:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
KW Nuclear protein.
SQ SEQUENCE 390 AA; 45286 MW; 39590CD3B61077CC CRC64;

Query Match 51.9%; Score 1057.5; DB 2; Length 390;
Best Local Similarity 50.7%; Pred. No. 2e-50;
Matches 207; Conservative 72; Mismatches 92; Indels 37; Gaps 7;

QY 1 MGKSDFLTPKATNRKSKGLKRLWYCMQCKQCDENGFKCHCSHQRQLLASEN 60
DB 1 MGRAEVGTPTKYLANKWKMSKGLKRLWYCMQCKQCDENGFKCHCSHQRQLLASEN 60
QY 61 PQOFMDYFSEERNDPFLRLRRFGTKRVHNNIVYNEYISHREHHIMNATQWETLTDFTK 120
DB 61 PGKFLHSFSEKSDGMYELLRRFRFGTKRTSANKIYQYIAHKEHHIMNATRWLTLSDYVK 120
```

```
QY 121 WLGRGLCKVDTPKGYIOYIDRDPETIRROLELEKXKQDLDEKTAKEIEBOVR-- 178
DB 121 WLGRTCQVIADETEKGFVTVYIDRSPEAMERQAKADKEMKEXKODDERWADFTIEQIKNA 180
QY 179 RGLGKGEQEB-VPTFTLSRENDEE-KVTFNLKSGACSSSGATSSKSTLGPALKITIGSS 236
DB 181 KAKDGEEDGQKFTLKRRENEPLKDIRLEK-----KFQPDVTLGKSAL----- 226
QY 237 ASVKRKESSQSTQSKKKKK-----SALDEIMEIEEEKK-RTARTDYWLQPE 284
DB 227 -----AKRPAPEAEKVKPKSVAGDSQTRSVDLDEIIKQESKKEKRAKRDYWLHKG 279
QY 285 IIVKIITKLGEYHKYKKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVTPAPCKRIL 344
DB 280 IIVKFTSKMGKFFKQKAVLVLDVDRYQKIKFLETGKLVQDQAHLETVPALDKPVM 339
QY 345 VLVGGYRGNEGTLSEINETSATIVETGPLKGRVVEGIQYEDISK 392
DB 340 VVNGAYRGSEALLKLDERRYSVSVLEIHLGPKLGRIVDNVQYEDISK 387

RESULT 9
Q9CV58 MOUSE
ID Q9CV58 MOUSE PRELIMINARY; PRT; 195 AA.
AC Q9CV58
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Mus musculus adult male tongue cDNA, RIKEN full-length enriched
DE library, clone:2310020017 product:antigenic determinant of rec-A
DE protein, full insert sequence. (fragment).
DE Name=Kin;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RA The FANTOM Consortium,
```

RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RL 60,770 full-length cDNAs.";
 RL Nature 420:563-573 (2002).
 [4]

RN NUCLEOTIDE SEQUENCE;
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RL Prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630 (2000).
 [5]

RN NUCLEOTIDE SEQUENCE;
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kiteunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RL sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771 (2000).
 [6]

RN NUCLEOTIDE SEQUENCE;
 RC STRAIN=C57BL/6J; TISSUE=Tongue;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito K., Saito R., Sakai K., Sakai K.,
 RA Sano H., Saeki D., Shibata K., Shibata Y., Shingawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Teijina Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AK009429; BAB26281.2; -; mRNA.

DR MGI:196676; Kin.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR007087; ZnF_C2H2.
 DR SMART; SM00355; ZnF_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 FT NON TER 195 195
 SQ SEQUENCE 195 AA; 23518 MW; 7B112E9CA820BFA6 CRC64;

Query Match 51.3%; Score 1044; DB 2; Length 195;
 Best Local Similarity 98.5%; Pred. No. 5e-50; Indels 0; Gaps 0;
 Matches 192; Conservative 1; Mismatches 2;

QY 1 MGKSDFLTPKAIANRIKSGKGLRWYCMQCKQCRDENGFKCHCMSESHORQLLAS 60
 DB 1 MGKSDFLSPKAIANRIKSGKGLRWYCMQCKQCRDENGFKCHCMSESHORQLLAS 60
 QY 61 PQQPMDFSEBFNDFLELLRRRFGTKRVHNNIVYNEIYSHREIHNNATOWETLTPTK 120
 DB 61 PQQPMDFSEBFNDFLELLRRRFGTKRVHNNIVYNEIYSHREIHNNATOWETLTPTK 120
 QY 121 WLGRGLCKVDTPKGYIOYIDRDPETIRQLKLEKKKQDLDEKTAKEIEQVRRG 180
 DB 121 WLGRGLCKVDTPKGYIOYIDRDPETIRQLKLEKKKQDLDEKTAKEIEQVRRG 180
 QY 181 LEGKEQVPTFTLS 195
 DB 181 LEGKEQVPTFTLS 195

RESULT 10

Q9ZVUS ARATH
 ID Q9ZVUS_ARATH PRELIMINARY; PRT; 411 AA.
 AC Q9ZVUS;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-PEB-2005 (TREMBLrel. 29, Last annotation update)
 DE TSA14.13 protein (Hypothetical protein At1g55460).
 GN Name-TSA14.13; Synonyms=At1g55460;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
 RA Altati H., Araujo R., Rowley D., Buehler E., Dunn P.,
 RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
 RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
 RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,
 RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 [3]
 RP NUCLEOTIDE SEQUENCE.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bower L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC005223; AAD10649.1; -; genomic_DNA.
 DR EMBL; AY051011; AAK93688.1; -; mRNA.
 DR EMBL; AF360132; AAK25842.1; -; mRNA.
 DR PIR; H96596; H96596.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR006646; KOW sub.
 DR InterPro; IPR007087; ZnF_C2H2.
 DR Pfam; PF00467; KOW; 1.
 DR SMART; SM00739; KOW; 1.
 DR SMART; SM00355; ZnF_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Hypothetical protein.

QY SEQUENCE 411 AA; 47288 MW; 9DA6F8648002065D CRC64;
 Query Match 48.0%; Score 978; DB 2; Length 411;
 Best Local Similarity 48.6%; Pred. No. 5e-46;
 Matches 201; Conservative 66; Mismatches 123; Indels 24; Gaps 6;

QY 1 MGKSDFLTPKAIANRIKSGKGLRWYCMQCKQCRDENGFKCHCMSESHORQLLAS 60
 DB 1 MGKSDFLTPKAIANRIKSGKGLRWYCMQCKQCRDENGFKCHCMSESHORQLLAS 60
 QY 61 PQQPMDFSEBFNDFLELLRRRFGTKRVHNNIVYNEIYSHREIHNNATOWETLTPTK 120
 DB 61 PTVVDGYSSEFEQTFDLARRSHRFRSIAATVYNEIYNDRHVHNNSTEWATLTFTK 120
 QY 121 WLGRGLCKVDTPKGYIOYIDRDPETIRQLKLEKKKQDLDEKTAKEIEQVRRG 180
 DB 121 WLGRGLCKVDTPKGYIOYIDRDPETIRQLKLEKKKQDLDEKTAKEIEQVRRG 180
 QY 181 LEGKEQVPTFTLS 195
 DB 181 LEGKEQVPTFTLS 195

```

QY 181 LE-----GKEQVPTFTLSRENDEB-----KVTNLSKGACS--SSGATSSK 221
DB 181 ABLKNGGGEGEGETSGNDEVDGDDDERKKDEDLRLKSGVKVGFALGGVQVATCKERGE 240
QY 222 STLGPSALKITGSSASVVRKSSOSSQTSQKSKKKSALDIMEIEEEK--RTARTDYW 280
DB 241 SSKLLFGDEENDKVRGEKKKSGDS--GRSEKERSALDLMKEEEKKKERMKRDYW 297
QY 281 LOPEIIVKIITKLGK-YHKKAIVKEVIDKYTAVVWMIDSGDKLKDQTHLETVPAP 339
DB 298 LPEGIIVKMSKALAEKGYKQGVKVIDNVGSEIKMLDSKHVLRVDPQKLETVLPQI 357
QY 340 GKRILVNGGYRGNEGTLESINEKTFSATIVETGPKRRRVEGTYQYEDISKLA 393
DB 358 GGMVIVNGAYGRGNARLLGVDTERFCAKVGIEKGVYDGRVKSIEYEDICKLA 411

RESULT 11
Q9XWF2_CABEL PRELIMINARY; PRT; 404 AA.
AC Q9XWF2;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Hypothetical protein Y52B11A.9.
GN ORFNames=Y52B11A.9;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916; DOI=10.1126/science.282.5396.2012;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL032654; CAA21720.1; -; Genomic_DNA.
DR PIR; T27106; T27106.
DR WormBase; WBGene00013128; Y52B11A.9.
DR WormPep; Y52B11A.9; CE20297.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0007275; P:development; IEA.
DR InterPro; IPR002358; Ribosomal_L6_1.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00525; RIBOSOMAL_L6_1; UNKNOWN 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 404 AA; 46401 MW; 9F55A1413525F057 CRC64;

Query Match 46.8%; Score 952.5; DB 2; Length 404;
Best Local Similarity 46.6%; Pred. No. 1.3e-44;
Matches 193; Conservative 80; Mismatches 110; Indels 31; Gaps 8;

QY 1 MCKSDFLTPTKATNRITKSGLOKLRWYCOMCQKQCDENGFKCHCNSHSHORQLLASEN 60
DB 1 MGKHEGSKDKLANRTKSGLOKLRWYCOMCQKQCDANGFKCHLTSEAHORQLLFAEN 60
QY 61 PQOFMDYFEEFRNDFLELRRRFGPKRVHNNIYVNEYISHREIHNNATOWETLDFPK 120
DB 61 SNSYLRFQSFNDKPNMQLRTSYGKRVANEVYNAPFKDGHVHMNSTVWHSLTGFVQ 120
QY 121 WLGREGLCVKDPTPKGWYIYQIDRDPETIRROLELEKKKKQDLDDDEKTAFTIEBOVRG 180
DB 121 YLGSSGCKXIDGDKGWYAYIDQ--EALIRKEEDQKQKQKDDBERHMQMDGMVQRG 178
QY 181 --LEGKEQVPTFTLSRENDEBKVTFNLSKGACS-----SSGATSSKSTL----- 225

```

```

DB 179 KELAGDDHEYEATELIRDTQDKIQLDLNLGILDRKLDVKSGVASAKISIFDMPKVKE 238
QY 226 -----GPSALKITIGSSASVVRKSSOSSQTSQKSKKKSALDIMEIEEEK--RTARTDY 279
DB 239 DPDEPGPS-----QPSRKSGKRSRSRSPAACKFKSKSALDEIKEMBEERKKERKNRKY 292
QY 280 WLOPEIIVKIITKLGKGYHKKKAIVKEVIDKYTAVVWMIDSGDKLKDQTHLETVPAP 339
DB 293 WREGIIVKVIITKLSGSEYRAKGVVRKVDVDTAQVAK-LDDGTVVVKLDQBEHVEVIFSL 351
QY 340 GKRILVNGGYRGNEGTLESINEKTFSATIVETGPKLRGRVEGTYQYEDISKLA 393
DB 352 GROMVINGARGQEAATLESIDEXFSLRLKIASPTGREGID-VPYEDASKLA 404

RESULT 12
Q4TDV9_TETNG PRELIMINARY; PRT; 269 AA.
AC Q4TDV9;
DT 13-SEP-2005 (TRENBLrel. 31, Created)
DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TRENBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF6005, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG0002601001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthonard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Bionnet C., Skalli Z., Cattolico L., Poullain J., De Berardinis V.,
RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAABE01006005; CAF88923.1; -; Genomic_DNA.
FT NON TER 1
FT NON TER 269
SQ SEQUENCE 269 AA; 31736 MW; E6883839A2CD17AF CRC64;

Query Match 46.0%; Score 936.5; DB 2; Length 269;
Best Local Similarity 69.8%; Pred. No. 6e-44;
Matches 178; Conservative 36; Mismatches 32; Indels 9; Gaps 4;

QY 55 LLASENPOOFMDYFEEFRNDFLELRRRFGPKRVHNNIYVNEYISHREIHNNATOWET 114
DB 1 LLASENPNRFDHFSQEFKDFLELRRRFGPKRVHNNIYVNEYISHREIHNNATRWET 60
QY 115 LTDFTKWLGREGLCVKDPTPKGWYIYQIDRDPETIRROLELEKKKKQDLDDDEKTAFTIE 174
DB 61 LTDFTKWLGREGLCVKDPTPKGWYIYQIDRDPETIRROLELARKKKHDLDDDEERSAKFTIE 120
QY 175 EQVRGREGKE-QEYPTFTLSRENDEBKVTFNLSKGACSSSGATSSKSTLGPALKTI 233

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:23:48 ; Search time 34.0867 Seconds
(without alignments)
953.202 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 2037
Sequence: 1 MGKSDFLPKAANRIKSG.....GPLKGRVEGIQYEDISKLA 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5 COMB.pdp.*
2: /cgn2_6/prodata/1/iaa/6 COMB.pdp.*
3: /cgn2_6/prodata/1/iaa/7 COMB.pdp.*
4: /cgn2_6/prodata/1/iaa/8 COMB.pdp.*
5: /cgn2_6/prodata/1/iaa/9 COMB.pdp.*
6: /cgn2_6/prodata/1/iaa/10 COMB.pdp.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	379.5	18.6	136	2	US-09-270-767-32008
2	360	17.7	274	2	US-09-248-796A-20041
3	132	6.5	1507	2	US-08-929-329-5
4	127	6.2	2349	2	US-09-538-092-914
5	123	6.0	1055	2	US-09-949-016-9776
6	122.5	6.0	1564	2	US-10-144-198-2
7	122.5	6.0	1564	2	US-10-144-198-4
8	122.5	6.0	2662	2	US-09-595-684B-31
9	122.5	6.0	2663	2	US-09-538-092-1252
10	122	6.0	934	2	US-09-949-016-6779
11	120.5	5.9	793	2	US-09-538-092-1271
12	120.5	5.9	1003	2	US-09-198-452A-17
13	120	5.9	440	2	US-09-762-194-2
14	120	5.9	440	2	US-09-762-194-4
15	120	5.9	586	1	US-08-630-822A-70
16	120	5.9	586	1	US-09-005-069-70
17	120	5.9	586	2	US-09-171-156A-30
18	120	5.9	586	2	US-09-004-730A-30
19	120	5.9	586	2	US-08-981-799A-30
20	120	5.9	1211	2	US-09-134-001C-4820
21	119.5	5.9	1007	2	US-09-155-558-30
22	118.5	5.8	425	2	US-09-286-981B-11
23	118.5	5.8	425	2	US-10-254-995-11
24	118.5	5.8	583	2	US-09-710-279-1358
25	118.5	5.8	671	2	US-09-439-313-380
26	118.5	5.8	671	2	US-09-352-616A-380
27	118.5	5.8	671	2	US-09-289-198-306

28	118.5	5.8	671	2	US-09-636-215-380	Sequence 380, App
29	118.5	5.8	671	2	US-09-685-166A-380	Sequence 380, App
30	118.5	5.8	671	2	US-09-429-755-306	Sequence 306, App
31	118.5	5.8	671	2	US-09-679-426-380	Sequence 380, App
32	118.5	5.8	671	2	US-09-759-143-380	Sequence 380, App
33	118.5	5.8	671	2	US-09-651-236-380	Sequence 380, App
34	118.5	5.8	671	2	US-09-699-295-306	Sequence 306, App
35	118.5	5.8	671	2	US-09-534-825A-306	Sequence 306, App
36	118.5	5.8	671	2	US-09-657-279-380	Sequence 380, App
37	118.5	5.8	671	2	US-10-012-896-380	Sequence 380, App
38	118.5	5.8	1003	2	US-09-438-185A-8	Sequence 8, Appli
39	118.5	5.8	1010	2	US-09-134-001C-5178	Sequence 5178, Ap
40	118.5	5.8	1719	2	US-09-439-313-378	Sequence 378, App
41	118.5	5.8	1719	2	US-09-352-616A-378	Sequence 378, App
42	118.5	5.8	1719	2	US-09-636-215-378	Sequence 378, App
43	118.5	5.8	1719	2	US-09-685-166A-378	Sequence 378, App
44	118.5	5.8	1719	2	US-09-679-426-378	Sequence 378, App
45	118.5	5.8	1719	2	US-09-759-143-378	Sequence 378, App

ALIGNMENTS

RESULT 1
US-09-270-767-32008
; Sequence 32008, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 32008
; LENGTH: 136
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-32008

Query Match 18.6%; Score 379.5; DB 2; Length 136;
Best Local Similarity 54.1%; Pred. No. 1.2e-26;
Matches 72; Conservative 28; Mismatches 32; Indels 1; Gaps 1;

Qy	261	LDIMEIEEKK-RTARTDYWLOPEIIIVKLTIKLGEKHYKKKAIYKEVIDKYTAVVKMI	319
Db	1	LDEIIKQESKKERANKDYWLHKGIVWKFISKMGKFFQKAVLVDVDRYQGRKIKFL	60
Qy	320	DSGDKLKLDQTHLETVIPAPGRILVNGGYRGNEGTLSEINKEKTSATIVETGPLKGR	379
Db	61	ETGEKUKVDQAHLETVIPALDKPVMVNGAYRSEALLKLDERRYSVSVILLHGLKGR	120
Qy	380	RVEGIQYEDISKL 392	
Db	121	IVDNVQYEDISKL 133	

RESULT 2
US-09-248-796A-20041
; Sequence 20041, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13

```
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 20041
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-20041

Query Match      17.7%; Score 360; DB 2; Length 274;
Best Local Similarity 32.9%; Pred. No. 1.9e-24;
Matches 91; Conservative 52; Mismatches 78; Indels 56; Gaps 9;

QY 1 MCKSDFLPKAIANIKSGKGLQKAWYCOMCKQCRDENGPKCHOMCSHQRQL--LLAS 58
Db 36 MAKAEFGTAKYQSKLRAAGLQKLPYCOLCKQCRDNGFKNHLSSPHIKKVSIEHES 95
QY 59 ENPOQFMDYFSEPRNDFLELLRRRFGTKRVHNNIVYNEYISHREHIHNNATQWETLTDF 118
Db 96 GDSSKLIETYSTKFDQKFIKLRIHNGTKFINANKFYQEIYIRERDHIHNNSTRWSLTSP 155
QY 119 TKWLREGGLCKV-----DETPKM--YIQYIDRDPETIRROLELEKXKKQDLDEEKT 170
Db 156 IKHLGKNGIVKQTNDESNEEBEGFNLEIKLIDRTQTLNAYQIKSGGNSLED-NDEMND 214
QY 171 KFIEQVRGLGKEQVPTFTFELSRNDEEKVTFNLSKGACSSGATSSKSTGLGPSAL 230
Db 215 KLLQKQIKRG-----QEMEKDKDAE-----KSKTLE----- 240
QY 231 KTIGSSASVKRKESQSSQSKK-----KKKKSALDE 263
Db 241 QLVPTSPVK-----LTLKKKKTATTCKLVNAFDE 270

RESULT 3
US-08-929-329-5
; Sequence 5, Application US/08929329
; Patent No. 6120770
; GENERAL INFORMATION:
; APPLICANT: Adams, John H
; APPLICANT: Dalton, John P
; APPLICANT: Kappe, Stefan
; TITLE OF INVENTION: Plasmodium Proteins Useful for Preparing
; TITLE OF INVENTION: Vaccine Compositions
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Barnes & Thornburg
; STREET: 11 S Meridian
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/929,329
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Breen, John P
; REGISTRATION NUMBER: 38,833
; REFERENCE/DOCKET NUMBER: 835910-28685
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 231-7745
; TELEFAX: (317) 231-7433
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1507 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein

; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium yoelii
US-08-929-329-5

Query Match      6.5%; Score 132; DB 2; Length 1507;
Best Local Similarity 19.5%; Pred. No. 0.012;
Matches 67; Conservative 59; Mismatches 131; Indels 86; Gaps 8;

QY 44 HCMSESHORQLLASENPOQFMDYFSEPRNDFLELLRRRFGTKRVHNNIVYNEYISHRE 103
Db 1046 HNSAPEHFRSLKINSYTPNRRGENFAKE-----SDSTNTDESKWDEVIRKKE 1093
QY 104 HIHNNATQWETLTDFTKWLREGGLCKVDETPKGWYIQYIDR-----PETIRROLE 154
Db 1094 EAAKNA-----EIIKRPPEAQKAAWAKKAAEERKKAEEKAAEAEERKRIE 1138
QY 155 LEKKKKQDLDEEKTAKFIEQVRGLGKEQVPTFTFELSRNDEEKVTFNLSKGACSS 214
Db 1139 AEKAAEERKRIEAKKAAEERKRIEAKKAAEERKRIEAKKAAEERKRIEAKKAAE 1198
QY 215 ----SGATSSKSTGLGPSALKTIGSSASVKRKESQSSQSKK-----KKKKSALDE 270
Db 1199 RKTEAAKAAEERKKAEEAVK--AEAAKKAEEAKKALKESRGRGKTTIEAVKKAEE 1256
QY 271 KKR-----TARTDYMLOPEIIVKIITKGLGKRYHKKKAIK 306
Db 1257 RKRIEAKKAAEERKRIEAVKKQKKKNVEAEKAAEERKRIEAAKAAEERIKREAVK 1316
QY 307 E-----VIDKYTAV-----VKMIDSGDKLKDQ 329
Db 1317 EEEVVIKKNLSEKKTAIFLKNSNNYETRNIDNNSPKKLDE 1359

RESULT 4
US-09-538-092-914
; Sequence 914, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 914
; LENGTH: 2349
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P12270
US-09-538-092-914

Query Match      6.2%; Score 127; DB 2; Length 2349;
Best Local Similarity 21.0%; Pred. No. 0.063;
Matches 69; Conservative 61; Mismatches 114; Indels 84; Gaps 13;

QY 48 ESHQRQLLASENPOQFMDYFSEPR--NDFL-----ELRRRFGTKRVH 90
Db 121 EAERKDLIRTNRLSQQELEYLTEDVKRLNEKLNKESNTTKGELQLKLDELQADSVKYRE 180
QY 91 NNIVYNEYISHREHIHNNATQWETLTDFTKWLREGGLCKVDETPKGWYIQYIDRDPETIR 150
Db 181 KLEQEKELHSHSQNTWLN--TELKTKTDELLALGRE-----KGNELDE 221
```


QY 151 RQLELEKKKKQ--DLDE-----EKTAKPIEEOVRGLEGKEQEVPTFTTSLRENDE 200
Db 222 LKCNLENKEEVRLEQWNGLKTSNEHLOKHVEDLTLTKAKEAQOASMEKPFHNLNA 281
QY 201 EKVTNLSKGCSSGATSSKSTLGPALKTIGSSASVKKRESSQSSQSKKKKSA 260
Db 282 HIKLSNLYKSAADSEAKSNELTRAVEELHKL-----KEAGEAN-----KAI 324
QY 261 LBIMEIEBKKTARTDWLQPEIIVKI--ITKLG-----EKYHKKAIKVE----- 307
Db 325 QHLLVEQSKDQ-----MEKEMLEKIGRLEKELENDLLSATKRKGAILEBELAA 377
QY 308 VIDKYTAVVMIDSGDKL-KLDQTHLET 334
Db 378 MSPTAAAVAKIVPGMKULTELYNAYVET 405

RESULT 5
US-09-949-016-9776
; Sequence 9776, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9776
; LENGTH: 1055
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9776

Query Match 6.0%; Score 123; DB 2; Length 1055;
Best Local Similarity 18.6%; Pred. No. 0.045;
Matches 85; Conservative 94; Mismatches 157; Indels 120; Gaps 21;

QY 18 SKGLQKLRYCQCKQ-----CRDENGFKCHCMSESHORQLLLASENPQFMDYF 68
Db 516 SKEKEDLOEKCIDIWEKLAQTKRVLAANEENS-----KMEQSNLEKLELNVKLOQLDQL 571
QY 69 SEE---FRNDFLELLRRRFGTKRVHNNVINYISHREHIHMN-----AT 110
Db 572 NRDKLSLHND- ISAMQOQLQEKREAVNSLOBELANVQDHLNLAQDLHTTKHQDVLLSE 630
QY 111 QWETLDTFTWLGBREGLCVKDETPKGYIYIDRDP-----TIRQLELEKKKKQDLD 165
Db 631 QTRLOKDISEANRFDCOKEETKQOQLQVLQNEISENKLKVQBSMMFQLOKERESE 690
QY 166 EEK--TAKFTEEOVRGLE-----GKEQEVPTFTTSLRENDEERKVTF 205
Db 691 ESKLETSKVLKEQHQHLEKLTDOQSKLODVLSKVLAAEERTVLOE-----EEWCE 744
QY 206 NLSKGCSSGATSSKSTLGPAA--LKTIGSSASVKKRESS-----QSSQSKKKKKKS 259
Db 745 SLENTLSQTKRQLSERSQQLVEKSGELLALQKEADSMRADFSLRNQFLTERKKAERQVA 804
QY 260 ALDIMEIE-----BEKKTARTDWLQPEI-IVKIITKLGKHYHKKKAIKVEI 309
Db 805 SLKEALKIORSQLEKNLLEOKQENS-----CIQEMATIELVAQ---DNHERARRLMKELN 857
QY 310 D---KYTAVVMKI-----DSGDKLKLD-----QTHLETV-----IPAPGK 341

Db 858 QMQYEYTELKQMANQKOLEROMRISDAMRTLKSEVKDEIRTSKLKLNOLFELPADLE 917
QY 342 RILVLNGGYRGEGTLESINERTFSGATIVETGPKL 377
Db 918 AILERN---ENLEGELESKE---NLPTMNEGPF 947

RESULT 6
US-10-144-198-2
; Sequence 2, Application US/10144198
; Patent No. 6833247
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-2

Query Match 6.0%; Score 122.5; DB 2; Length 1564;
Best Local Similarity 24.2%; Pred. No. 0.089;
Matches 64; Conservative 37; Mismatches 87; Indels 77; Gaps 12;

QY 116 TPTFKWLGREGLCVKDET-----PKG-WYIQYIDRDPETIRROL-ELEKKKKQDLD 165
Db 40 TQCRNWL-KTGNCLYGNTCFVHGPPSPRGKGYSSNYRRSPERTGDLRRMRMKNKRODVT 98
QY 166 E-----EKTAKFIEHQVRGLEGKEQEVPTFTTSLRENDEEKVTFNLSKGACS----- 213
Db 99 EPQKRNTTESSSPVRKESRGRHREKDIKITKERTPESEENVEWETNRDDSDGDINY 158
QY 214 -----SSGATSSKSTLGPALKTIG 234
Db 159 DYVHELSELMKQKIQRELKMLEQENKEXKEEIIKKKVSPEVVRSKLSP-SFSLRK--- 214
QY 235 SSASVVRKRESSQSSQSKKKKKKSAALDEIME-----IEEEKKTARTDWLQPEIIVK 288
Db 215 SSKSPKRKSPKSSSSAKDKRTSAVSSPLLDQORNSKTNQSKKGGPRT--SPPPPIPE 272
QY 289 IITKLGKHYHKKKAIKVEIDKYT 313
Db 273 DIA--LGKKY-KEKYVKORIEBK 294

RESULT 7
US-10-144-198-4
; Sequence 4, Application US/10144198
; Patent No. 6833247
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-4

Query Match 6.0%; Score 122.5; DB 2; Length 1564;
Best Local Similarity 24.2%; Pred. No. 0.089;
Matches 64; Conservative 37; Mismatches 87; Indels 77; Gaps 12;

QY 116 TPTFKWLGREGLCVKDET-----PKG-WYIQYIDRDPETIRROL-ELEKKKKQDLD 165

```
Db 40 TCRNLW-KTGNCLVNTCRFHVGSFRGKGSSNVRSPERPTGDLRMRMKNKQDVDT 98
QY 166 E-----EKTAKFIEBQVARGLEGKEQVPTFTLSRENDEKVTFNLSKGACS----- 213
Db 99 EPQKNTESSPVRKSSGRHREKEDIKITRTPESSEENVEWETNRDDSDNGDINY 158
QY 214 -----SSGATSSKSTLSPSALKTTG 234
Db 159 DYVHLSLEMKRQKIQRELMLKEQENMEKREBIIIKKEYSPEVVRSKLSP-SPSLRK---- 214
QY 235 SSASVKRKESSOSTOSKEKKKKSALDEIMB-----IEBEKRTARTDYWLQPEIIVK 288
Db 215 SKSPARKSPKSSGASKDKRTSAVSSPLDQQRNKTNQSKKGPRTP--SPPPPIPE 272
QY 289 IITKLGKEYHKKAIVKEVIDKYT 313
Db 273 DIA--LGKKY-KEKYVKDRIEKT 294

RESULT 8
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

Query Match 6.0%; Score 122.5; DB 2; Length 2662;
Best Local Similarity 21.2%; Pred. No. 0.2;
Matches 81; Conservative 67; Mismatches 159; Indels 75; Gaps 14;

QY 17 KSKGLQKLWYCOMQKQCRDENGFKCHMS-ESHQROLLLASENPQQPMDYFSEFRND 75
Db 1104 KHAHKEGELSRTCRLAEVEEKLKESQQLQEQQLLVQVEEMSEMQKINE----- 1158
QY 76 FLELLRRRFGTKRVHNNIVYNEYISHREHIMNATOWETLDTFKWLGREGCKVDETPK 135
Db 1159 -IENLKNELKNKELTLEHETERLELAQKLNENYEVKSITKERVYL--KELQSFETER 1215
QY 136 G---WYIQYIDRDPETIRROLE---LEKKKKQDLDE-----EKTAKFIEQ-VRRGLE 182
Db 1216 DHLRGYIREIATGQTKEELKIAHILKHEQETIDELRRSVSEKTAQIINTQDLEKSH 1275
QY 183 GKEQVPTFTLS-----RENDEKVTFN-----LSKGACSSSGATSS----- 220
Db 1276 KLQEBIPVLHEEQELLPNVKVSETQETMNELELLTEQSTTKDSTTLARIEMERLRLNEK 1335
QY 221 -----KSSTLGPSALKTIGSSASVKR-----KSSQSTOSQSKKKKK 258
Db 1336 FOESQEEIKSLTKERDNLTKITKEALEVGHQDLKEHIRETLAKIQESQSQKQSLNKKED 1395
QY 259 SALDEIM-EIEBEKRTARTDYWLQPEIIVKIITKLGKEYHKKAIVKEVIDKYTAVVK 317
Db 1396 NETTKIVSEMEQPKDSAL---LRIEIMGLSKRLQESHDEMSKVAKEKDDQLRLQEV 1452
```

```
QY 318 MIDSGDKLK-----LDQTHLET 334
Db 1453 LQESDQLKENIKETIVAKHLET 1474

RESULT 9
US-09-538-092-1252
; Sequence 1252, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurPatSeqFormatter Version 0.9
; SEQ ID NO 1252
; LENGTH: 2663
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)-(0)
; OTHER INFORMATION: Polypeptide Accession Number Q02224
US-09-538-092-1252

Query Match 6.0%; Score 122.5; DB 2; Length 2663;
Best Local Similarity 21.2%; Pred. No. 0.2;
Matches 81; Conservative 67; Mismatches 159; Indels 75; Gaps 14;

QY 17 KSKGLQKLWYCOMQKQCRDENGFKCHMS-ESHQROLLLASENPQQPMDYFSEFRND 75
Db 1105 KHAHKEGELSRTCRLAEVEEKLKESQQLQEQQLLVQVEEMSEMQKINE----- 1159
QY 76 FLELLRRRFGTKRVHNNIVYNEYISHREHIMNATOWETLDTFKWLGREGCKVDETPK 135
Db 1160 -IENLKNELKNKELTLEHETERLELAQKLNENYEVKSITKERVYL--KELQSFETER 1216
QY 136 G---WYIQYIDRDPETIRROLE---LEKKKKQDLDE-----EKTAKFIEQ-VRRGLE 182
Db 1217 DHLRGYIREIATGQTKEELKIAHILKHEQETIDELRRSVSEKTAQIINTQDLEKSH 1276
QY 183 GKEQVPTFTLS-----RENDEKVTFN-----LSKGACSSSGATSS----- 220
Db 1277 KLQEBIPVLHEEQELLPNVKVSETQETMNELELLTEQSTTKDSTTLARIEMERLRLNEK 1336
QY 221 -----KSSTLGPSALKTIGSSASVKR-----KSSQSTOSQSKKKKK 258
Db 1337 FOESQEEIKSLTKERDNLTKITKEALEVGHQDLKEHIRETLAKIQESQSQKQSLNKKED 1396
QY 259 SALDEIM-EIEBEKRTARTDYWLQPEIIVKIITKLGKEYHKKAIVKEVIDKYTAVVK 317
Db 1397 NETTKIVSEMEQPKDSAL---LRIEIMGLSKRLQESHDEMSKVAKEKDDQLRLQEV 1453
QY 318 MIDSGDKLK-----LDQTHLET 334
Db 1454 LQESDQLKENIKETIVAKHLET 1475

RESULT 10
US-09-949-016-6779
; Sequence 6779, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
```

;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;; FILE REFERENCE: CL001307
;; CURRENT APPLICATION NUMBER: US/09/949,016
;; CURRENT FILING DATE: 2000-04-14
;; PRIOR APPLICATION NUMBER: 60/241,755
;; PRIOR FILING DATE: 2000-10-20
;; PRIOR APPLICATION NUMBER: 60/237,768
;; PRIOR FILING DATE: 2000-10-03
;; PRIOR APPLICATION NUMBER: 60/231,498
;; PRIOR FILING DATE: 2000-09-08
;; NUMBER OF SEQ ID NOS: 207012
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 6779
;; LENGTH: 994
;; TYPE: PRT
;; ORGANISM: Human
US-09-949-016-6779

Query Match 6.0%; Score 122; DB 2; Length 994;
Best Local Similarity 18.6%; Pred. No. 0.051;
Matches 85; Conservative 94; Mismatches 157; Indels 120; Gaps 21;
QY 18 SKGLOKLRYCQMCQKQ-----CRDENGFKCHMSHSHORQLLASNPQQFMDYF 68
DB 455 SKEKEDLQKCDIWEKKLAQTKRVLAABENS-----KMEQSNLEKLDLNVKRLQQLDQL 510
QY 69 SEB---FRNDLLELLRRRFGTKRVHNNIVVNEYISHREHHM-----AT 110
DB 511 NRDLKSLHND-ISAMQOQLQEKREAVNSLQELANVQDHLNLAQKQDLHTTKHQDVLLSE 569
QY 111 QWETLTDFTKWLREGCLCKVDETPKGYIYIDRDP-----TIRQLELEKKKKODLDD 165
DB 570 QTRLQKDI SEWANRFEDCQBEETKQOQLOVLQNEIEENKLVUQOEMMFQRLQKERESE 629
QY 166 EEK--TAKFTIEQVRGLE-----GREQEYPTFTLSRENDESKVTF 205
DB 630 ESKLETSKYTLKEQOQLEKELTDQSKLDQVLSKVLAABERVTLQE-----EERWCE 683
QY 206 NLSKGACSSGATSKSSTLGPAA--LKTIGSSASVKRKESS-----QSSTQSEKKKKK 259
DB 684 SLEKTLSTQKRLSREQQVLVEKSGELLALQKEADSNRADFSLLRNQFLTERKKAQKQVA 743
QY 260 ALDEIMEIE-----EKKGTARTDYWLQPEI-IVKIITKLGEKVKHKKAIIVKEVI 309
DB 744 SLKEALKIORSQLEKNLLEQKQNS---CQKEMATIELVAQ---DNHERARRLMKELN 796
QY 310 D---KYTAVVVMi-----DSGDKLKLDD-----QTHLETV-----IPAPGK 341
DB 797 QMQUEYTELKQMANQKDLERQWEISDAMRTLKSEVKDEIRTSLNKLNQFLPELPADLE 856
QY 342 RILVLNGYVGRNGTLESINEKTSFATIVITGPKL 377
DB 857 AILERN---ENLEGELESLEK---NLPFTWNEGPFE 886

RESULT 11
US-09-538-092-1271
; Sequence 1271, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormat Version 0.9
; SEQ ID NO 1271

;; LENGTH: 793
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q05682
US-09-538-092-1271
Query Match 5.9%; Score 120.5; DB 2; Length 793;
Best Local Similarity 17.6%; Pred. No. 0.05;
Matches 65; Conservative 57; Mismatches 125; Indels 123; Gaps 9;
QY 48 ESHQQLLASNPQQFMDYFSEPRNDFLELLRRRFGTKRVHNNIVVNEYISHREHHM 107
DB 105 ERRQKRLQELARQKEFQPTITDASLS-----LPSSRMQNDTAENETTEKEKSES 155
QY 108 NATQWETLTDFTKWLREGCLCKVDETPKGYIYIDRDPETIRRLQ-----ELEKKKKQ 161
DB 156 RQERVE-----IET-----ETVTYSYQKNDWRDAEENKKE 186
QY 162 DLDDEKTAFTIEQVRGLEKQEVPTFTLSRENDESKVTFNLSKGCSSSGATSSK 221
DB 187 DKEKEEEE---EKPKRGSIGENQVVMVVEBKTTESQBEETVMSLKNQOISSEEPKQEE 242
QY 222 SSTLGPAL-----KTIGSSAS----- 238
DB 243 EREQSDSEISHHEKMEEDKERAERARLEAEERERIKAEQDKKIADERARIEAEKAA 302
QY 239 -----VKRESSOSTQSEKKEKKKKKKKALDIMEIEEE 270
DB 303 AQERERRAEERERMRKEEKRAAEERQRIKEBEKRAAEERQRIKEBEKRAAEERQRIKEE 362
QY 271 KRTA--RTDYWLQPEIIVKIITKLGEKVKHKKAIIVKEVIDKYTAVVMIDSG--DKLK 326
DB 363 EKRAAEERQRAAEBEKAKVBEQKRNKQLEBEKKRAMQETKINGEKVQEKGVNEKK 422
QY 327 LDQTHLETVI 336
DB 423 AQEDKLQTAIV 432
RESULT 12
US-09-198-452A-17
; Sequence 17, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention of infection and treatment of infection
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 17
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...1003
; OTHER INFORMATION: Xaa-unknown or other
US-09-198-452A-17
Query Match 5.9%; Score 120.5; DB 2; Length 1003;
Best Local Similarity 20.9%; Pred. No. 0.071;
Matches 81; Conservative 55; Mismatches 155; Indels 97; Gaps 16;
QY 5 DFLTPKAIAIRIKSGLOKRLWTCQMCQKQ--CRDENGFKCH-----CMSHSHORQLLAS 59
DB 466 DSIFKKIIDN-----FEKLAWKFMILSKSICRFTIIFENHGHVAKSLHKNVLEK- 518

```
QY 60 NPOQFMDYFSEFRNDFLELLRRRFGTKVHNNIVY-----NEYISHREHHMNATQWETL 115
Db 519 -----VIYRSLQKSYRDIGMSSAKMKILHGNPFFSLEDNKKTKIMKEHAEM-----LES 567
QY 116 TDFTK---WLGREGCKVDTEPKG-----YIQ 140
Db 568 SSYRKVFALSDENVVDFSPDKKDWLSDGIPCRDALSEISRDEQWQKKAHLKHQESLYTQ 627
QY 141 YIDR--DPETIRRLQLEKKKQDLDDEKTAFTIEEQVRRGLEKGEQVPTFTLSRPN 198
Db 628 ARDLTDQSSKENQKAEQYISSWVRKKFEIERVQERIQAIQKLYPNILEREET 687
QY 199 DEEKVTNLSKACSSSGNT-----SSKSTGLPSALKTIGSSASVKRKSSQSSQTQ 250
Db 688 TQETVTPTVQGTASSDLDTLGRIEVSSREDNQNE-----SCVKVLRSHVEMSGE 741
QY 251 SKEK--KKKKSALDEIMETEEBKRTARTDYWLQPEIIVKIITKKLGEKYHKKKAIVKEV 308
Db 742 VKQEGPKKKEFDQMGSLR-----FFTEHIEEVLQKDYSKHL--SYFKKVNKKEV 794
QY 309 IDKYTAVVMIDSGDKLKDQTHLETVI 336
Db 795 --QYAKF-----RLKVLESDLEGIL 812

RESULT 13
US-09-762-194-2
; Sequence 2, Application US/09762194
; Patent No. 6835539
; GENERAL INFORMATION:
; APPLICANT: Elbaz, Nathalie
; APPLICANT: Nahmias, Clara
; APPLICANT: Strosberg, Arthur Donny
; TITLE OF INVENTION: NUCLEIC SEQUENCES ENCODING AN AT2
; FILE OF INVENTION: RECEPTOR-INTERACTING PROTEIN (ATIP) AND THEIR APPLICATIONS
; FILE REFERENCE: 33339/208804
; CURRENT FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: PCT/FR99/01908
; PRIOR FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: FR 98/09997
; PRIOR FILING DATE: 1998-08-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-762-194-2

Query Match 5.9%; Score 120; DB 2; Length 440;
Best Local Similarity 20.8%; Pred. No. 0.023;
Matches 66; Conservative 64; Mismatches 129; Indels 58; Gaps 11;

QY 41 FKCHCMSES----HORQLLASENPQPMDFSEFRNDFLELLRRRFGTKVHNNIVYN 96
Db 74 YTKCESQSGFILHLRQLLSRGNKFEALTIVIQHLSEREALQKQ-----HKTLSQ 125
QY 97 EYISHREHHMNATQWETLDTFTKWLREGLCCKVD-ETPKGWYIQYIDRDPETIRRLQEL 155
Db 126 ELVSLRGELVAASSACEKLEK-----ARADLTAYQEFVQKLNQHQHQTDRTELEN 175
QY 156 EKKKKQDLDDEKTAFTIEEQVRRGLEKGEQVPTFTLSRNDDEKVTNLS-----208
Db 176 RLKDYLTAECKLQSIYIEAEKYTQLQEQ----FDNLNAAHETTKLEIEASHSEKVEL 231
QY 209 -KGACSSSGATSSKSTGLPSALKTIGSSASVKRKSSQSSQSKKKKKKKSALDEIMEI 267
Db 232 LKTYTETSLSEIKKSHMEKKSLEDL-----LNEKQESLEK-QINDLKSNDALNERLKS 285
QY 268 EEEKRTARTDYWLQPEIIT-----VKIITKKLGEKYHK---KKAIVKEVIDKYTAVV 316
Db 286 EEQKQLSREKANSKPNQVMYLEQESLKAIVLEIKNEKLHQQDMKLMKMEKLVNDNNTALV 345

QY 97 EYISHREHHMNATQWETLDTFTKWLREGLCCKVD-ETPKGWYIQYIDRDPETIRRLQEL 155
Db 126 ELVSLRGELVAASSACEKLEK-----ARADLTAYQEFVQKLNQHQHQTDRTELEN 175
QY 156 EKKKKQDLDDEKTAFTIEEQVRRGLEKGEQVPTFTLSRNDDEKVTNLS-----208
Db 176 RLKDYLTAECKLQSIYIEAEKYTQLQEQ----FDNLNAAHETTKLEIEASHSEKVEL 231
QY 209 -KGACSSSGATSSKSTGLPSALKTIGSSASVKRKSSQSSQSKKKKKKKSALDEIMEI 267
Db 232 LKTYTETSLSEIKKSHMEKKSLEDL-----LNEKQESLEK-QINDLKSNDALNERLKS 285
QY 268 EEEKRTARTDYWLQPEIIT-----VKIITKKLGEKYHK---KKAIVKEVIDKYTAVV 316
Db 286 EEQKQLSREKANSKPNQVMYLEQESLKAIVLEIKNEKLHQQDMKLMKMEKLVNDNNTALV 345
```

```
QY 317 KVIDSGDKLKDQTHLE 333
Db 346 -----DKLKRFPQENE 356

RESULT 14
US-09-762-194-4
; Sequence 4, Application US/09762194
; Patent No. 6835539
; GENERAL INFORMATION:
; APPLICANT: Elbaz, Nathalie
; APPLICANT: Nahmias, Clara
; APPLICANT: Strosberg, Arthur Donny
; TITLE OF INVENTION: NUCLEIC SEQUENCES ENCODING AN AT2
; FILE OF INVENTION: RECEPTOR-INTERACTING PROTEIN (ATIP) AND THEIR APPLICATIONS
; FILE REFERENCE: 33339/208804
; CURRENT APPLICATION NUMBER: US/09/762,194
; CURRENT FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: PCT/FR99/01908
; PRIOR FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: FR 98/09997
; PRIOR FILING DATE: 1998-08-04
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-762-194-4

Query Match 5.9%; Score 120; DB 2; Length 440;
Best Local Similarity 20.8%; Pred. No. 0.023;
Matches 66; Conservative 64; Mismatches 129; Indels 58; Gaps 11;

QY 41 FKCHCMSES----HORQLLASENPQPMDFSEFRNDFLELLRRRFGTKVHNNIVYN 96
Db 74 YTKCESQSGFILHLRQLLSRGNKFEALTIVIQHLSEREALQKQ-----HKTLSQ 125
QY 97 EYISHREHHMNATQWETLDTFTKWLREGLCCKVD-ETPKGWYIQYIDRDPETIRRLQEL 155
Db 126 ELVSLRGELVAASSACEKLEK-----ARADLTAYQEFVQKLNQHQHQTDRTELEN 175
QY 156 EKKKKQDLDDEKTAFTIEEQVRRGLEKGEQVPTFTLSRNDDEKVTNLS-----208
Db 176 RLKDYLTAECKLQSIYIEAEKYTQLQEQ----FDNLNAAHETTKLEIEASHSEKVEL 231
QY 209 -KGACSSSGATSSKSTGLPSALKTIGSSASVKRKSSQSSQSKKKKKKKSALDEIMEI 267
Db 232 LKTYTETSLSEIKKSHMEKKSLEDL-----LNEKQESLEK-QINDLKSNDALNERLKS 285
QY 268 EEEKRTARTDYWLQPEIIT-----VKIITKKLGEKYHK---KKAIVKEVIDKYTAVV 316
Db 286 EEQKQLSREKANSKPNQVMYLEQESLKAIVLEIKNEKLHQQDMKLMKMEKLVNDNNTALV 345

QY 317 KVIDSGDKLKDQTHLE 333
Db 346 -----DKLKRFPQENE 356

RESULT 15
US-08-630-822A-70
; Sequence 70, Application US/08630822A
; Patent No. 5840695
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LYNDIA
; TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
; FILE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
```

STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,822A
FILING DATE: 11-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 586 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Xaa = any amino acid
LOCATION: 379
US-08-630-822A-70

Query Match 5.9%; Score 120; DB 1; Length 586;
Best Local Similarity 21.4%; Pred. No. 0.036;
Matches 63; Conservative 61; Mismatches 115; Indels 56; Gaps 12;
Qy 139 IQYIDRDPETIRROLELEKKKKQDLDDDEEKTAKFIEQVRRLGEGKEQEVPTFTLSREN 198
Db 305 VEFTKKDLEVVEDALDTLSKDKNNLVIEKEVIKDIKEEIIADYQEDVEELKEAIVAAEKPK 364
Qy 199 DEEKVTNLSKGACSSGATSSKSTLGSALKTIGSSASVYRK-----ESSQSSTQSKE 253
Db 365 DEIKET---KGA-QRLKKVKNMTRWDITVQVIESKESEKAKTLPLEAPRSATETQE 419
Qy 254 -----KKKKKSGALDEIMEIEEKGKT-----ARTDYWLQPEIIVKLI-- 290
Db 420 LDVRKERGEIUIIDELMDAIIKKVKNVPDENRLKLIENILGRIDTDKDRHIKVEDVLYKVIDI 479
Qy 291 -----TKKLG--KYHKKAIV--KEVIDKYTAVVK-MIDSGDKLKLDOHLETVI 336
Db 480 VEKEDGIMSTKQDLVQLKKEVIELEKEKEQESQQSFPVPSFTLHLESSQOKSTV 539
Qy 337 PAPGKRILVNLGGYRGNGLTESINEKTFSATIVETGPL-KGRRVEGIQYEDIS 390
Db 540 PSSGHEAKV-----SEDDL-NVKNKLEESTKTECGAIDEEHREHCQYDPDIT 586

Search completed: November 23, 2005, 16:35:43
Job time : 36.0867 secs

This Page Blank (used)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:24:04 ; Search time 114.291 Seconds
(without alignments)
1436.746 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 2037
Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPLKGRVRGQIYEDISKLA 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap:*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap:*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap:*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap:*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap:*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2037	100.0	393	4	US-10-408-765A-1465
2	1059.5	52.0	390	6	US-11-097-143-41040
3	934	45.9	437	4	US-10-425-114-48183
4	934	45.9	437	4	US-10-425-114-48183
5	929	45.6	424	4	US-10-425-115-341788
6	909.5	44.6	423	5	US-10-739-930-8499
7	617	30.3	176	4	US-10-424-599-168326
8	535	26.3	138	4	US-10-437-963-191916
9	453	22.2	120	4	US-10-767-701-57151
10	370	18.2	84	4	US-10-106-698-5464
11	326	16.0	162	4	US-10-424-599-196166
12	239	11.7	123	4	US-10-425-115-300949
13	222.5	10.9	107	4	US-10-425-115-230934
14	209	10.3	109	4	US-10-424-599-196168
15	150.5	7.4	278	4	US-10-425-114-46572
16	147.5	7.2	430	4	US-10-425-115-284883
17	145.5	7.1	1114	4	US-10-408-765A-2119
18	145.5	7.1	1213	5	US-10-788-793-6
19	144.5	7.1	1212	5	US-10-788-793-2
20	144	7.1	375	5	US-10-739-930-6249
21	138	6.8	805	4	US-10-369-493-152
22	136	6.7	624	4	US-10-105-959-4
23	136	6.7	624	4	US-10-258-662-20
24	135.5	6.7	512	4	US-10-108-260A-4080
25	135.5	6.7	921	5	US-10-732-923-3305
26	133	6.5	1359	5	US-10-732-923-8707
27	133	6.5	1359	5	US-10-732-923-8708

ALIGNMENTS

RESULT 1

US-10-408-765A-1465
; Sequence 1465, Application US/10408765A
; Publication No. US20040101874A1

GENERAL INFORMATION:

; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1465
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1465

Query Match 100.0%; Score 2037; DB 4; Length 393;

Best Local Similarity 100.0%; Pred. No. 2.8e-143;

Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGKSDFLTPKAIANRIKSGKGLQKLRWYCMQCKQCRDENGPKCHCMESHORQLLAS	60
Db	1	MGKSDFLTPKAIANRIKSGKGLQKLRWYCMQCKQCRDENGPKCHCMESHORQLLAS	60
Qy	61	PQQFMDYFSEFRNDFLELLRRRGTGKRVNNIYVYSHREIHNNATOWELTDFTK	120
Db	61	PQQFMDYFSEFRNDFLELLRRRGTGKRVNNIYVYSHREIHNNATOWELTDFTK	120
Qy	121	WLGREGKCKVDTPKGYIQYIDRDPETIRRLQLEKKKKODLDEEKTAKFIEQVRRG	180
Db	121	WLGREGKCKVDTPKGYIQYIDRDPETIRRLQLEKKKKODLDEEKTAKFIEQVRRG	180
Qy	181	LEGKEQVPTTELGRNDEBEKVTFNLSKGACSSSGATSSKSTLGP9ALKTISSASVK	240
Db	181	LEGKEQVPTTELGRNDEBEKVTFNLSKGACSSSGATSSKSTLGP9ALKTISSASVK	240
Qy	241	RKSSQSSQSTQSEKSKKKSALDEIMEIEEKKRTARTDYWLQPEIIVKIITKLGK	300
Db	241	RKSSQSSQSTQSEKSKKKSALDEIMEIEEKKRTARTDYWLQPEIIVKIITKLGK	300
Qy	301	KKAIIVKIDKYTAVVWKMDSGDKLKDQTHLETVPAPGKRILVNLNGYRGNEGTL	360

||||| 301 KKAIVKEVDKTTAVVWMDSDGKLDQTHLETVIPAPGKRLVUNGIRNGEGLTSLI 360
Db
||||| 361 NEKTSATIVETGPKLGRVREGIOYEDISKLA 393
Qy
||||| 361 NEKTSATIVETGPKLGRVREGIOYEDISKLA 393
Db
RESULT 2
US-11-097-143-41040
; Sequence 41040, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41040
; LENGTH: 390
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-41040

Query Match 52.0%; Score 1059.5; DB 6; Length 390;
Best Local Similarity 50.7%; Pred. No. 1.7e-70;
Matches 207; Conservative 73; Mismatches 91; Indels 37; Gaps 7;
Qy 1 MGKSDFLTPKAIANRIKSKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGRAEVGTPKYLANKMKSKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHORQLLASEN 60
Qy 61 PQOFMDYFSEEPNDPFLRLRRFGRKRVHNNIVVNEYSIASHREHIHNNATOWETLTDFTK 120
Db 61 PGKFLHSFKESFGSDGFWELLRRRFRGKRTSANKIYQEIYAHKEHIHNNATRLWLTSDYVK 120
Qy 121 WLGRGLCKVDTPKGYIYQIDRDPETIRRLQLELEKKKKQDDEEKTAKFTIEQVR-- 178
Db 121 WLGRGTVIADETEKGMFWYIYIDRSPAMERQAKADKEKMEKDDDEERMADEFTIEQIKNA 180
Qy 179 RGLGKQEB--VPTFTLSRENDEE-KVTNLSKGACSSGATSSKSTGLPSALKTIGSS 236
Db 181 KAKDGEDEGQEKFTBLKRENEPLKDIRLEK-----KFPDPTVLGKSAL----- 226
Qy 237 ASVKRESQSSQTSQSEKKKK-----SALDEIMEIEEKK-RTARTDYLQPE 284
Db 227 -----AKRPAPAESEKVFPPKPSVAGSGQTSVLDEIIKQESSEKERNRKYWLHG 279
Qy 285 IIVKIITKGLKHYHKKAIKVEVDKYTAVVWMDSDGKLDQTHLETVIPAPGKRL 344
Db 280 IIVKFTSKMGEKFFPKQAVLVLDIVDYOQKIKFLETGKGLKYDQAHLETVIPALDKPM 339

Qy 345 VLVGGRYRNEGTLSINSEKTSATIVETGPKLGRVREGIOYEDISKL 392
Db 340 VVNGAYRGEALLRKLDERRYSVVEILHGLPKGRIVDNVQVEDISKL 387
RESULT 3
US-10-425-114-48183
; Sequence 48183, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48183
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI.pep
US-10-425-114-48183

Query Match 45.9%; Score 934; DB 4; Length 437;
Best Local Similarity 45.3%; Pred. No. 4.4e-61;
Matches 193; Conservative 75; Mismatches 122; Indels 36; Gaps 7;
Qy 1 MGKSDFLTPKAIANRIKSKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHORQLLASEN 60
Db 14 MGKHEFLTPKAIANRIKAKGLQKLRYWCQMCQKQCRDENGFKCHCMSESHORQMVFGMA 73
Qy 61 PQOFMDYFSEEPNDPFLRLRRFGRKRVHNNIVVNEYSIASHREHIHNNATOWETLTDFTK 120
Db 74 PDVVVEGFSBFLSFLSLIRRAHRSVAATVVVYVNEIADRRHHVHNNSTRWATLTFEYK 133
Qy 121 WLGRGLCKVDTPKGYIYQIDRDPETIRRLQLELEKKKKQDDEEKTAKFTIEQVR-- 179
Db 134 FLGREGYCKVEDTPKGMFWYIYDRSEQAVKRLKRRKIKSDMADDERQERMIARQIERA 193
Qy 180 -----GLEGKEQEVPTFTLSRENDEEKTFTNLSKGACSSG--ATSSKSTLGP 227
Db 194 HKS LAKPNGGGAAGEPESGSEGESGSDDDDEQEPEDDSKEADKATGKIAIALQKAVPCP 253
Qy 228 SALKTIGSASVKRK-----ESSQSSTQSEKKK-----KKSALDEIMEIEE 269
Db 254 ---KVPFPDDPKPMKFGFDEEDSGTRDQKNELTKKMGKDVKAAAEAKRSALDELMKEE 310
Qy 270 E-KKTARTDWLQPEIIVKIITKLGK-KYHKKAIKVEVDKYTAVVWMDSDGKLD 327
Db 311 KAKERSNRKDYWLCGIVVVKWSKSLAEKGYIYQKGVVKVYIDKYVGETEMESKHVLRV 370
Qy 328 DQTHLETVIPGKRLVUNGIRNGEGLTSLINSEKTSATIVETGPKLGRVREGIOY 387
Db 371 DQDELETVIPQGLVLRVUNGAYRGSNARLLSVDTKEKCAKQVQEKGLYDGVLAIVEY 430
Qy 388 DISKLA 393
Db 431 DICKIS 436
RESULT 4
US-10-425-114-58982
; Sequence 58982, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong

APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 58982
LENGTH: 437
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: 700238628_FLI.pep
US-10-425-114-58982

Query Match 45.9%; Score 934; DB 4; Length 437;
Best Local Similarity 45.3%; Pred. No. 4.4e-61;
Matches 193; Conservative 75; Mismatches 122; Indels 36; Gaps 7;
Qy 1 MGKSDFLTPKAIANRIKSGLOKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 14 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFGMA 73
Qy 61 PQOFMDYFSEFRNDPFLRLRRFGTKRVHNNIVYNEYISHRHHIHMNATOWETLTDTFK 120
Db 74 PDRVGFSEFLESFLSLIRRAHRSRVAATVVYNEYIADRHVHMNSTRWATLTFFVK 133
Qy 121 WLREGCLKVDTPKGYIOYIDRDPTIRROLELEKKKKQDLDDEKTAFTIEEQVRR- 179
Db 134 FLGREGYCKVEDTPKGFMTYIDRDSEQAVKRLKRIKSDMADDERQERMIARQIERA 193
Qy 180 -----GLEKGEQVPTFTLSRENDEEKVTNLSKGACSSG--ATSSKSSTLGP 227
Db 194 HKSLAKPNGGGAEGEPESGEYSDDDEQEDDSKEADKATGKIAIAIQKAVPGP 253
Qy 228 SALKTTIGSSASVVRK-----ESSOSSTQSKEKKK-----KKSALDEIMEIEE 269
Db 254 ---KVPFDDKPKMGFGDEEDSGTRDQEKNELTKMGKDVKAEEAKRSALDELMEKEE 310
Qy 270 E-KKRTARTDYWLOPEIIVKIITKLGEK-YHKKAIVKEVIDKYTAVVMIDSGDKKL 327
Db 311 KAKERSNRKDYWLCPGIIVKVMKSLAEKGYKGVVKKVVDKYVGEIEMLESKHVLRV 370
Qy 328 DQTHLETVIPAPKRIILVNGYRGNEGTLSEINETSATIVETGPKGRRVEGIOYE 387
Db 371 DQDELETVIPQIGGLVRVNGAYRGSNARLLSVDTKFCVAKVQVEKGLYDGKVLRAVEYE 430
Qy 388 DISKLA 393
Db 431 DICKIS 436

RESULT 5
US-10-425-115-341788
Sequence 341788, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
Plants
FILE REFERENCE: 38-21(53322)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 341788
LENGTH: 424

TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_74879C.1.pep
US-10-425-115-341788

Query Match 45.6%; Score 929; DB 4; Length 424;
Best Local Similarity 45.3%; Pred. No. 1e-60;
Matches 193; Conservative 73; Mismatches 124; Indels 36; Gaps 7;
Qy 1 MGKSDFLTPKAIANRIKSGLOKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGKHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFGMA 60
Qy 61 PQOFMDYFSEFRNDPFLRLRRFGTKRVHNNIVYNEYISHRHHIHMNATOWETLTDTFK 120
Db 61 PDRVGFSEFLESFLSLIRRAHRSRVAATVVYNEYIADRHVHMNSTRWATLTFFVK 120
Qy 121 WLREGCLKVDTPKGYIOYIDRDPTIRROLELEKKKKQDLDDEKTAFTIEEQVRR- 179
Db 121 LLGREGYCKVEDTPKGFMTYIDRDSEQAVKRLKRIKSDMADDERQERMIARQIERA 180
Qy 180 -----GLEKGEQVPTFTLSRENDEEKVTNLSKGACSSG--ATSSKSSTLGP 227
Db 181 HKSLAKPNGGGAEGEPESGEYSDDDEQEDDSKEADKATGKIAIAIQKAVPGP 240
Qy 228 SALKTTIGSSASVVRK-----ESSOSSTQSKEKKK-----KKSALDEIMEIEE 269
Db 241 ---KVPFDDKPKMGFGDEEDSGTRDQEKNELTKMGKDVKAEEAKRSALDELMEKEE 297
Qy 270 E-KKRTARTDYWLOPEIIVKIITKLGEK-YHKKAIVKEVIDKYTAVVMIDSGDKKL 327
Db 298 KAKERSNRKDYWLCPGIIVKVMKSLAEKGYKGVVKKVVDKYVGEIEMLESKHVLRV 357
Qy 328 DQTHLETVIPAPKRIILVNGYRGNEGTLSEINETSATIVETGPKGRRVEGIOYE 387
Db 358 DQDELETVIPQIGGLVRVNGAYRGSNARLLSVDTKFCVAKVQVEKGLYDGKVLRAVEYE 417
Qy 388 DISKLA 393
Db 418 DICKIS 423

RESULT 6
US-10-739-930-8499
Sequence 8499, Application US/10739930
Publication No. US20040216190A1
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
FILE REFERENCE: 38-21(53377)B
CURRENT APPLICATION NUMBER: US/10/739,930
CURRENT FILING DATE: 2003-12-18
NUMBER OF SEQ ID NOS: 11088
SEQ ID NO 8499
LENGTH: 423
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C7583_2.p
US-10-739-930-8499

Query Match 44.6%; Score 909.5; DB 5; Length 423;
Best Local Similarity 44.6%; Pred. No. 2.8e-59;
Matches 189; Conservative 76; Mismatches 126; Indels 33; Gaps 8;
Qy 1 MGKSDFLTPKAIANRIKSGLOKLRWYCMQCKQCRDENGFKCHCMSESHORQLLASEN 60
Db 1 MGRHEFLTPKAIANRIKAKGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQVFGMA 60
Qy 61 PQOFMDYFSEFRNDPFLRLRRFGTKRVHNNIVYNEYISHRHHIHMNATOWETLTDTFK 120

```
Db 61 PDRVVEGFSEFLESFLIRRAHRSRVAATVVYNEYIADRRHHVHMNSTRWATLTFVK 120
Qy 121 WLGRGLCKVDTPKGWYIYIDRDPETIRRQLELEKKKKQDLDDDEKTAKEFBEQVRR- 179
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 121 LLRGEGYCKVEDTPKGWFWYIDRSEQAVKDLKRIKSDMAEDERQERMIARQIERA 180
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Qy 180 -----GLEGEQVPTFTELSRNDEBEKVTFNLSKGACSSG--ATSSKSSTLG 226
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 181 HKSLAKPNDGDAEG-EPESGGEYSGDDDELEEDVSKANKATGKIAIALQTAVRG 239
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Qy 227 PSALATIGSSASVKRKESSQTSQKEKK----KKK-----SALDIMEETE-E 270
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 240 PK-VNPFEDKPEVKRFPDEGLGTGEEKDEFLAKGKGVKAAADARRSALDELMKEEMA 298
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Qy 271 KKTARTDWLQPELIIKTIITKLGEK-YHKKAIVKEVIDKYTAVVKMIDSGDKLKLDQ 329
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 299 KERSNRKDWLCPGIVKVMKSLAEKGYKQGVVRKVMKYVGEIEMLESKHVLRVDQ 358
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Qy 330 THLETVIPAGKRILVNLGGYRGNEGTLSEINEKTFSATIVITGTPLKRRRVEGIQYEDI 389
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 359 DELETVIPQIGLVRILNGAYRGSNARLLSDVTEKFCAKVQVEKGLYDGKVLRAVKYEDI 418
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Qy 390 SKLA 393
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 419 CKIS 422
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|

RESULT 7
US-10-424-599-168326
; Sequence 168326, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 168326
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1.pep
US-10-424-599-168326

Query Match 30.3%; Score 617; DB 4; Length 176;
Best Local Similarity 64.9%; Pred. No. 5.5e-38;
Matches 109; Conservative 25; Mismatches 34; Indels 0; Gaps 0;

Qy 1 MGKSDFLTPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHQRLLLASEN 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 1 MGKNEFLTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSEGHQRMQIFGN 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Qy 61 PQOFMDYFSEFRNDFLELRRRFGTKRVHNNIVYNEYISHREHIMNATOWETLTDFTK 120
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 61 PRIVEGYSEEFSTPLEMKKSHRSRVAATVVYNEYINDRRHHIMNSTQWATLTFVK 120
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Qy 121 WLGRGLCKVDTPKGWYIYIDRDPETIRRQLELEKKKKQDLDDDEK 168
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 121 YLGRTGCKVETPTKGFWYIDRDSSETLTKERMKNKRIKADWDVEK 168
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|

RESULT 8
US-10-437-963-191916
; Sequence 191916, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

```
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 191916
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1.pep
US-10-437-963-191916

Query Match 26.3%; Score 535; DB 4; Length 138;
Best Local Similarity 67.4%; Pred. No. 5.2e-32;
Matches 93; Conservative 21; Mismatches 24; Indels 0; Gaps 0;

Qy 1 MGKSDFLTPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHQRLLLASEN 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 1 MGKHEFLTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSESHQRMQVFGQA 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Qy 61 PQOFMDYFSEFRNDFLELRRRFGTKRVHNNIVYNEYISHREHIMNATOWETLTDFTK 120
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 61 PDRVVEGFSEFLDAFLTLRRRAHRSRIAATVTVNEFIADRRHHVHMNSTRWATLTFVK 120
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Qy 121 WLGRGLCKVDTPKGWY 138
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 121 FLGRSGHCKVEDTPKGF 138
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|

RESULT 9
US-10-767-701-57151
; Sequence 57151, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 57151
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30963818.pep
US-10-767-701-57151

Query Match 22.2%; Score 453; DB 4; Length 120;
Best Local Similarity 67.5%; Pred. No. 5.5e-26;
Matches 81; Conservative 14; Mismatches 25; Indels 0; Gaps 0;

Qy 1 MGKSDFLTPKAIANRIKSGLOKLRYWCQKQCRDENGFKCHCMSESHQRLLLASEN 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 1 MGKHEFLTPKAIANRIKAKGLQKLRYWCQKQCRDENGFKCHCMSESHQRMQVFGMA 60
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Qy 61 PQOFMDYFSEFRNDFLELRRRFGTKRVHNNIVYNEYISHREHIMNATOWETLTDFTK 120
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
Db 61 PDRVVEGFSEFLESFLSLIRRAHRSRVAATVVYNEYIADRRHHVHMNSTRWATLTFVK 120
|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:|
```

```
RESULT 10
US-10-106-698-5464
; Sequence 5464, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5464
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5464

Query Match      18.2%; Score 370; DB 4; Length 84;
Best Local Similarity 98.6%; Pred. No. 5.3e-20;
Matches 73; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 320 DSGDKLKLDTHTLETVPAPGKRILVNGYRGNEGTLSEINKEKTSATIVETGPKGR 379
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 11 ESGDKLKLDTHTLETVPAPGKRILVNGYRGNEGTLSEINKEKTSATIVETGPKGR 70

QY 380 RVEGIQVEDISKLA 393
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 71 RVEGIQVEDISKLA 84

RESULT 11
US-10-424-599-196166
; Sequence 196166, Application US/10424599
; Publication No. US2004031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196166
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19163C.1.pep
US-10-424-599-196166

Query Match      16.0%; Score 326; DB 4; Length 162;
Best Local Similarity 43.7%; Pred. No. 2.3e-16;
Matches 69; Conservative 32; Mismatches 45; Indels 12; Gaps 3;

QY 238 SVKEKSSQSTQSKKKKKSALDEIMEIEEEKK-RTARTDYWLQPEIIVKIITKKLGE 296
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 15 NKKKESGGG-----KSALDEMMREEEKKKEKINRKYWLHGEIVVVKWSKVLAE 64

QY 297 K-YHKKAIVKEVDIKYAVVKMIDSGDKLKDTHLETVPAPGKRILVNGYRGNEG 355
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 KGYKQKGVKRVKIDKYVGEIEMLESKHLRVDAQLELTVPQVGGRVKIVNGAYRGSA 124

QY 356 TLESINEKTSATIVETGPKGRRVEGIQVEDISKLA 393
```

```
Db 125 KLLGVDTDNFCAKQVIEKGAYDGRVLKSVYEYEDICKVA 162

RESULT 12
US-10-425-115-300949
; Sequence 300949, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 300949
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(123)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37535C.1.pep
US-10-425-115-300949

Query Match      11.7%; Score 239; DB 4; Length 123;
Best Local Similarity 46.5%; Pred. No. 4.9e-10;
Matches 46; Conservative 14; Mismatches 39; Indels 0; Gaps 0;

QY 10 KAIANRIKSKGLQKLRYCOMCQKCRDENGFKCHCMSESHQOLLIASENPOQFMDYFS 69
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 25 KAIANRIKXRRGLQKLQXYCHMCHRRDQKELKCHCMRETHQRNMHVLGMSTDRVVGFS 84

QY 70 BEFRNDFLELLRRRFGTKRVHNNIVVNEYISHREHIMN 108
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 85 KEFLXSFLSFIRHADRHSRLTATXVYNYKYIVDRYHVMN 123

RESULT 13
US-10-425-115-230934
; Sequence 230934, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 230934
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_142205C.1.pep
US-10-425-115-230934

Query Match      10.9%; Score 222.5; DB 4; Length 107;
Best Local Similarity 39.6%; Pred. No. 7e-09;
Matches 42; Conservative 29; Mismatches 34; Indels 1; Gaps 1;

QY 289 IITKKLGEK-YHKKAIVKEVDIKYAVVKMIDSGDKLKDTHLETVPAPGKRILVNL 347
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```

Db      1  VMKSLAEKYYKQGVRRKMPKQYGEIEMLSKHLVRVDQDELETVPQIGLLRLN  60

Qy      348  GGYRGNEGTLSESINEKTFSAITVIETGPLKRRVEGIQYEDISKLA  393
         |||:  |:::  |:::  |:::  |:::  |:::  |:::  |:::
Db      61  GAYRGSNARLLSVDTDEKFCQVQEKGLYDGYKVLRAVEYEDICKIS  106

RESULT 14
US-10-424-599-196168
; Sequence 196168, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 196168
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(109)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_19165C.1.pep
US-10-424-599-196168

```

[illegible]

```

RESULT 15
US-10-425-114-46572
; Sequence 46572, Application US/10425114
; Publication No, US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46572
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700442544_FLI.pep
US-10-425-114-46572
Query Match          7.4%   Score 150.5;   DB 4;   Length 278;

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:28:34 ; Search time 5.01276 Seconds
(without alignments)
237.628 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 2037
Sequence: 1 MKSDPLTPKAIANKRIKSG.....GPLKGRVEGIQVEDISKLA 393

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 17545 seqs, 3030971 residues

Total number of hits satisfying chosen parameters: 17545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New.*
1: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB_PEP.*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB_PEP.*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB_PEP.*
4: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB_PEP.*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB_PEP.*
6: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB_PEP.*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB_PEP.*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118.5	5.8	583	1	US-10-793-626-1358
2	111.5	5.5	885	1	US-10-793-626-1660
3	106.5	5.2	2897	1	US-10-499-715-2
4	106	5.2	1189	7	US-11-074-176-134
5	104.5	5.1	1448	1	US-10-485-517-212
6	101.5	5.0	5024	1	US-10-793-626-2964
7	101	5.0	674	1	US-10-507-275-9
8	100	4.9	1299	1	US-10-821-234-1145
9	99	4.9	745	1	US-10-793-626-1742
10	98.5	4.8	752	1	US-10-793-626-348
11	95.5	4.7	676	7	US-11-135-855-28
12	95.5	4.7	717	7	US-11-135-855-29
13	95	4.7	422	1	US-10-821-234-1313
14	95	4.7	1155	1	US-10-793-626-1780
15	94.5	4.6	763	1	US-10-821-234-1619
16	94	4.6	1432	1	US-10-510-386-218
17	93	4.6	989	1	US-10-821-234-975
18	92.5	4.5	853	1	US-10-821-234-1110
19	91.5	4.5	299	1	US-10-793-626-1888
20	91.5	4.5	636	1	US-10-485-517-170
21	91.5	4.5	1279	1	US-10-793-626-3188
22	91	4.5	574	1	US-10-507-275-7
23	90.5	4.4	296	1	US-10-131-826A-26
24	90.5	4.4	672	1	US-10-689-742-70
25	90.5	4.4	702	1	US-10-510-386-214

26	90	4.4	410	1	US-10-793-626-3258	Sequence 3258, Ap
27	90	4.4	477	1	US-10-793-626-3250	Sequence 3250, Ap
28	90	4.4	919	1	US-10-821-234-1144	Sequence 1144, Ap
29	89.5	4.4	208	1	US-10-793-626-694	Sequence 694, App
30	89.5	4.4	208	1	US-10-793-626-1326	Sequence 1326, Ap
31	89	4.4	756	7	US-11-074-176-202	Sequence 202, App
32	88	4.3	368	1	US-10-689-742-100	Sequence 100, App
33	87.5	4.3	284	1	US-10-821-234-1632	Sequence 1632, Ap
34	87.5	4.3	594	1	US-10-131-826A-10	Sequence 10, Appli
35	87.5	4.3	654	7	US-11-046-668-4	Sequence 4, Appli
36	87.5	4.3	674	1	US-10-501-039-10	Sequence 10, Appli
37	87.5	4.3	745	1	US-10-793-626-1500	Sequence 1500, Ap
38	87	4.3	709	7	US-11-074-176-158	Sequence 158, App
39	87	4.3	747	1	US-10-821-234-1662	Sequence 1662, Ap
40	87	4.3	770	1	US-10-982-545-15	Sequence 38, Appl
41	87	4.3	770	1	US-10-789-273-38	Sequence 38, Appl
42	87	4.3	1586	1	US-10-821-234-901	Sequence 901, App
43	86.5	4.2	352	1	US-10-793-626-216	Sequence 216, App
44	86.5	4.2	1126	1	US-10-485-517-248	Sequence 248, App
45	86	4.2	752	1	US-10-793-626-1138	Sequence 1138, Ap

ALIGNMENTS

RESULT 1
US-10-793-626-1358
; Sequence 1358, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P034803US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1358
; TYPE: PRT
; LENGTH: 583
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1358

Query Match	5.8%	Score 118.5;	DB 1;	Length 583;
Best Local Similarity	20.4%	Pred. No. 0.075;		
Matches	87;	Conservative	75;	Mismatches 170; Indels 95; Gaps 17;
Qy	46	MSESHQRQLLASENPOQPMDF--SERFR-----NDFLELLRRRFGTKRVH--NNI 93		
Db	120	VNQNGQFIVQLLGVNAEQFRLFPQGEFKFLQSNKDKQILRLTFNSERFDEIRHL 179		
Qy	94	VYNEYISHREIHMNATOWETLDTFTKWLREGCLCKVDETPKGMVQIYIDRDE-----147		
Db	180	LVENVKQEKVQIENRYTQIENLWNDIDTFNNDELALYKELESSQTDKMIKFFQFNDYGC 239		
Qy	148	TIRROLEKKK-KQDLDD-----EETAKFIEQVRRLGEGKEQE-----187		
Db	240	KILKSFEAKNKITKELDDLNLHKYKYNVELSENTKKLKAERKIPDDLKKEQNYDKLQOE 299		
Qy	188	-----VPTFTTEL-SRENDEEKVTFNLSKGCSSGATSSKSTLSPALKTTGSS 236		
Db	300	LKMIQSSKVLITVTRLSLKKDKDELV-----SLHEQSKLNETNYHNEIKGPKQK 350		
Qy	237	-----ASVKRESSQSSQSKKKKKSALDEIME-----IEEKKRTARTDYLWQPEII 286		
Db	351	LEHLSTRENEITQFNOYLEKNQVFFNQDLKIISYQKQPVIEEIKKL-----YSEYNDLI 406		
Qy	287	VK--IITKLGKVKHKKKAIVKEVIDKYTAVVMIDSG-----DKLKLDQTHLET 334		

; Sequence 2964, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480U5
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2964
; LENGTH: 5024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic amino acid sequence
; OTHER INFORMATION: amino acid sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (5024)
; OTHER INFORMATION: variable amino acid

US-10-793-626-2964

Query Match 5.0%; Score 101.5; DB 1; Length 5024;
Best Local Similarity 20.7%; Pred. No. 18;
Matches 67; Conservative 60; Mismatches 119; Indels 77; Gaps 16;

QY 91 NNIVNYEYTSRH-----IHNATQTWETLTD--TKWLGRGLCK-VDETPKGWIY 139
Db 3463 NLHGQOKLAHAKODAAVINGLIHLNVAQREVMTNTATTREKYNALDNA----- 3516

QY 140 QYDRDPETRIRQLERLEKKKDLDBEE---KTAFIEBQVRGLEGEKEEV----- 188
Db 3517 QALDKAMETLU---QQVAHKNNILNDISKYLNEYSKY-QOOYDRIADAEQLLNQTTNP TL 3572

QY 189 -PPTFELSREN--DEKVTFNLGKGCSSSGATS-----SKSSTLGPSALKTIGSSASVK 240
Db 3573 EPKVDIVKDVLANEKILFGAEKLSDKSANDEIKHMVNLNAQAOSIKDMISHAALR 3632

QY 241 RKESQSSTQSKKKKKSLADEIMEETESEKTRATDYWLQPRLII VKTIITKKLGEKYHK 300
Db 3633 -----TBVKLLQOAKTLDEAMKLSLEDTQVVITDTL-----PNYTEASED 3674

QY 301 KKAIVKEVIDKYTA VVMKMDSGBLKLDQTH--LETVIPAGKRILVNGGYRGNEG T-- 356
Db 3675 KKEKVDQTVSHAQAIIDKI-NGSNVSLDVQRALQEQLTQASEN----LDGDORVEEKAVH 3729

QY 357 -----LESINEKTFSATI 369
Db 3730 ANQTIDLTHLSNQOTAKEVS 3752

RESULT 7
US-10-507-275-9
Sequence 9, Application US/10507275
Publication No. US20050250166A1
GENERAL INFORMATION:
APPLICANT: Masai, Hideo
APPLICANT: Tamai, Katsuyuki
APPLICANT: Medical and Biological Laboratories Co., Ltd.
APPLICANT: Japan Science and Technology Agency
APPLICANT: Ginkgo Biomedical Research Institute Co., Ltd.
TITLE OF INVENTION: Cdc7-ASK Kinase Complex, Substrates of the Kinase Complex,
TITLE OF INVENTION: Specific Antibodies to the Substrates, and Screening Methods
TITLE OF INVENTION: Using the Same to Screen for Compounds Comprising Cdc7-ASK
TITLE OF INVENTION: Kinase Inhibitory Ability
FILE REFERENCE: 082368-001100US
CURRENT APPLICATION NUMBER: US/10/507,275
CURRENT FILING DATE: 2004-09-09
PRIOR APPLICATION NUMBER: JP 2002-067702
PRIOR FILING DATE: 2002-03-12

```
; PRIOR APPLICATION NUMBER: WO PCT/JP03/02918
; PRIOR FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-507-275-9

Query Match
  Best Local Similarity 5.0%; Score 101; DB 1; Length 674;
  Matches 57; Conservative 51; Mismatches 102; Indels 80; Gaps 11;

QY 11 ATANRIKSLGKLRLWYQWQKQCRDENGFKCHMSHSHQROLLLASENPQQFMDYFSE 70
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 282 SIQLQLKEK---KKGYCECLQKYED---LETHLLSEQHRN---FAQSNQYQVDDIVS 332
QY 71 EFRNDPFLLELLRRRFGTKRVHNNIVYNEYISHREHIHMNATQWETLDTFTKWLREGLCVK 130
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 333 KLVDFFEVEKPTPKKRIKYSVGSLSPPVS-----ASVLKK 368

QY 131 DETPKGWYIYIDR-----DPETIRRLQLEKKKKQDLDEEKTAKFIEQV-----RR 179
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 369 TEQKEVELQHSQKDCQEDDTVKEQNFYK---ETQETKLLFISEPIPHPSNELR 424
QY 180 GLEGK-----EQEV-PTFTELSRNDEBEKVTFLNSKGACSSGATSKSSTLGP 227
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 425 GLNEKMSKCSMLSTAEDDIRQNTQLPLHKNQECILDISE-----HTLSE 471

QY 228 SALKTI-----GSSASVKRKESOSSQTSQKSKKKKKSALDEIMEIESE 270
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 472 NDLEELRVHYKCNIOASVHVSDFTDNGSQPKQKSDTVLPFAKDLKSK 521

RESULT 8
US-10-821-234-1145
; Sequence 1145, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1145
; LENGTH: 1299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1145

Query Match
  Best Local Similarity 4.9%; Score 100; DB 1; Length 1299;
  Matches 87; Conservative 68; Mismatches 144; Indels 170; Gaps 20;

QY 35 CRDE-----NGFKCHMSHSHQROLLLASENPQQFMDYFSEFRND-- 75
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 201 CRDNTSVYHISGKKTKFDVGNLLRSHGIDLHNRFLILQGEVEQIAMKPKQTEHDEG 260
QY 76 FLELLRRRFGTKRVHNNI-----VYNEYISHR-EHIHMNATQWETLDTFTKWLREG 126
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 MLEYLEDITGCGRLNEPIKVLCCRVRILNEHRGDKLNRVQWVEKEKDAL----- 309
QY 127 LCKVDSTPKGWYIQYIDRDPETIRQ-----LELEKK----- 158
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 310 -----EGEKNIALFELTLENEIFRKNHVQYIYELQKRIABMETQKEKIHDTKINE 364

; PRIOR APPLICATION NUMBER: US/10793626
; Sequence 1742, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1742
; LENGTH: 745
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1742

Query Match
  Best Local Similarity 4.9%; Score 99; DB 1; Length 745;
  Matches 66; Conservative 72; Mismatches 109; Indels 138; Gaps 15;

QY 33 KQCRDENGFKCHMSHSHQROLLLASENPQQFMDYFSEFRNDFLELLRRRFGTKRVHNN 92
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 KQREQNKIK-----KEEKAQLKEA-----KIERKKQKSRQNNN 186
QY 93 IVYNEYISHREHIHMNATQWETLDTFTKWLREGLCVKDETPKGWYIOVIDRDPETIRQ 152
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 187 VI-----KDVSDPFE-----ISQDDIDPIYGHNEGEDKRPNT----- 218
QY 153 LELEKKKKQDLDEEKTAKFIEEQVRRGLEKQEVPTFTELSRNDEKEKVTFF-----N 206
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 219 --ANQQRKVLNDEQ-----FQQLSPSTKQSNINNNQPSPTAENNQQS 260
QY 207 LSKGACSSGATSKSSTLGPSSALKTIGSSASVKRKESOSSQTSQKSKKKKKSALDEIM- 265
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 QAEGSISEAGEANETIYTPP-----LSLLKQPTKQKTTSKAEVQKQGVLESTLTK 311
QY 266 -----ETEEBKRTARTDYMLQPSIIIVKIIITKLGEKYHKKKAIIVKEVIDKYTAVVKMI 319
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 NFGVNAKVTQIKGPAVYQYEIQPAGVK-----VSKIVNLHNDIALAL 355
QY 320 DSGDKLKLDTQHTLETVP-----APGKRILVNGGYRGNEGTLSEINKEFTSATIVI 371
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 356 AAKD-----VRIEAPIGRSAVGIEVPNDKISLV-----TLKEVLEDKFPSPSKYL 400
```



```
Qy 372 ETGLKGRVVEG-----IQYEDISKL 392
Db 401 EVG--IGRDISGDPISQLNEMPHL 423

RESULT 10
US-10-793-626-348
; Sequence 348, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 348
; LENGTH: 752
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-348

Query Match 4.8%; Score 98.5; DB 1; Length 752;
Best Local Similarity 20.9%; Pred. No. 2.7;
Matches 71; Conservative 69; Mismatches 112; Indels 87; Gaps 19;

Qy 123 GREGCKVDPEPKGYI--QVIDRDP-ETIRROL-----ELEK----KKQDLDEEK 168
Db 174 GIEGKAYETGKGVVRSRDEPLRSGRKQLIVTEIPYEVNKSILVKRIDSLRADKK 233
Qy 169 TAKFIE---EQVRGLE-----GKEQEVPTTFLSRENDEEKVTFLNSKGCSSSGATSS 220
Db 234 VDGIVEVRDEYDRTGLRIABELKKDANSEIKNYLYKNSDLQISYFNFWAIS-----EG 288
Qy 221 KSSTLGPALTKTIGSSASVKKSSQSST---OSKEKK-----KKKSALDEIMIE 268
Db 289 RPKLMG---LREILSYLNHQIEVVTNRTYDLQAEKRMHVEGLMKALSILDEVIALI 345
Qy 269 EEKXR-----TARTDYW-LQPEIIVKI-----IT-----KKLGEYHKHKAIVKE---V 308
Db 346 RNSKNKKDKNLVAEYDTEAQAEI VMLQLYRLTNTDIEALKKEHELEALIKELRNI 405
Qy 309 IDKYTAVVKMI-----DSGDKLKLDOTHL-----ETVIPAPGKRILVINGGYR 351
Db 406 LDNHEALLAVIKDELNEIKKPKVDRLSTIEABISEIKIDKEVMVPSSEVILSLTQHG- 464
Qy 352 GNEGTLSEINKEPTGATIVETGPKLGRVVEGIVQVEDIS 390
Db 465 -----IKRTSRSFNASGVTEIGLXGDRLL--LKHSYV 496

RESULT 11
US-11-135-855-28
; Sequence 28, Application US/11135855
; Publication No. US20050255557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-29

Qy 144 RDETTIRROLELEKKKKQD-----LDDEETAKTAFIEQVRR-----GLEGK 184
Db 359 RRDEARRRELEARRRREQSEELRLREQK-----EEKERRRRADRGEARSGSGSD 413
Qy 185 EQEVPPTFTELSRENDEEKVTFLNSKGCSSSGATSSKSTLGPALKTKTIGSSASVKRKE 244
Db 414 E-----LREDDP-----VKGRGKGRGPPSSSDSEPEALELREAKKSAKKPQS 459
Qy 245 QSSTQSKEKKKKKSALDEIMEIEEKKRTARTDYWLOPEIIVKI-ITKKLGEKY----- 298

; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 676
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-28

Query Match 4.7%; Score 95.5; DB 7; Length 717;
Best Local Similarity 24.3%; Pred. No. 4.1;
Matches 51; Conservative 30; Mismatches 64; Indels 65; Gaps 10;

Qy 144 RDETTIRROLELEKKKKQD-----LDDEETAKTAFIEQVRR-----GLEGK 184
Db 359 RRDEARRRELEARRRREQSEELRLREQK-----EEKERRRRADRGEARSGSGSD 413
Qy 185 EQEVPPTFTELSRENDEEKVTFLNSKGCSSSGATSSKSTLGPALKTKTIGSSASVKRKE 244
Db 414 E-----LREDDP-----VKGRGKGRGPPSSSDSEPEALELREAKKSAKKPQS 459
Qy 245 QSSTQSKEKKKKKSALDEIMEIEEKKRTARTDYWLOPEIIVKI-ITKKLGEKY----- 298

; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-29
```

[illegible]

```
Qy 169 TAKFIERQ-----VRRGLEKQEVPFTFELSRENDEEKVT----- 204
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
264 GEEVVEDRDYVDTFKGDDYNEENPTEPGSDGTMSDKETHDVKAVCSQEAHTGPCRAVM 323
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : :
205 ----FNLSKGAC-----SSSGATSSKSTLGPSALKTIIGSSASVKKRES 244
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
324 PRWYFDLSKGKCVRFIYGGCGNRNPFESDYCMVAVCKAMIPPTPLPT--NDVDVYPETS 381
Qy : : : : : : : : : : : : : : : : : : : : : : : : : : : :
245 SQSSTQSKEKKKKKKSALDEIMEIEEKKRTARTDYWLQPEIIVKIITKKLGEKVHKKKAI 304
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
382 ADDNEHARFQAK-----EQLEIHRNRMDRVKKEWEEAELOAK-----NLPKAE 426
Qy 305 VKEVIDKYTAVMKID---SGDKLKLDQTHLETV 335
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
427 RQTLIQHFQAMVKALEKEAASEKQQLVETHLARV 460
```

Search completed: November 23, 2005, 16:39:52
Job time : 6.01276 secs

This Page Blank (usp1c)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:33:44 ; Search time 133.339 Seconds
(without alignments)
1295.011 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 393
Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPLKGRVVEGIQYEDISKLA 393

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 2443163 seqs, 439378781 residues

Word size : 0
Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A Geneseq 21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004as.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	393	100.0	393	2	AAY23620 A human k
2	393	100.0	393	7	ADJ69659 Human hea
3	167	42.5	391	2	AAY23619 Murine ki
4	165	42.0	293	2	AAY23618 A huamn t
5	120	30.5	291	2	AAY23617 Mouse tru
6	73	18.6	84	4	AAG74690 Human col
7	42	10.7	390	8	ADP22452 Sea-squir
8	35	8.9	411	3	AGG42576 Arabidops
9	35	8.9	423	3	AAB03064 Maize KIN
10	35	8.9	423	3	AAB03065 Maize KIN
11	35	8.9	423	8	ADT58422 Plant pol
12	35	8.9	424	3	AAB03063 Maize KIN
13	35	8.9	437	8	ADX96318 Plant ful
14	35	8.9	437	8	ADX78817 Plant ful
15	24	6.1	382	3	AGG42577 Arabidops
16	21	5.3	302	3	AGG46712 Arabidops
17	21	5.3	340	3	AGG46711 Arabidops
18	21	5.3	343	3	AGG46710 Arabidops
19	18	4.6	34	5	ADK35903 Novel hum
20	18	4.6	38	5	ADK35871 Novel hum
21	15	3.8	15	2	AAR66765 ReCa-like
22	15	3.8	390	4	ABB71416 Drosophil
23	11	2.8	11	2	AAR66766 Zinc-fing
24	8	2.0	8	5	ABG96595 Human leu

25	8	2.0	8	5	ABG97218 Human leu
26	8	2.0	8	7	ADL99055 Human leu
27	8	2.0	8	7	ADL98437 Human leu
28	8	2.0	55	3	AAG02279 Human sec
29	8	2.0	61	2	AAG75013 Human sec
30	8	2.0	61	5	ABG95475 Human nov
31	8	2.0	61	6	ABO34669 Region of
32	8	2.0	61	7	AD123330 Novel hum
33	8	2.0	61	8	ADH74332 Human sec
34	8	2.0	96	5	ABP01806 Human ORF
35	8	2.0	113	3	AAG54587 Zee maya
36	8	2.0	119	4	AAQ11954 Human pol
37	8	2.0	129	8	ADY13398 Plant ful
38	8	2.0	146	4	ABG15188 Novel hum
39	8	2.0	173	4	AAU58581 Propionib
40	8	2.0	173	6	ABM55100 Propionib
41	8	2.0	176	2	AAR98153 Thermosta
42	8	2.0	178	4	AAB96405 Putative
43	8	2.0	330	3	AAG11841 Arabidops
44	8	2.0	330	3	AAG39710 Arabidops
45	8	2.0	351	3	AAG48481 Arabidops

ALIGNMENTS

RESULT 1
AAY23620
ID AAY23620 standard; protein; 393 AA.
XX
AC AAY23620;
XX
DT 07-SEP-1999 (first entry)
XX
DE A human kin17 protein.
XX
KW Human; kin17 protein; cell proliferation; fertility;
KW hyperproliferative disease; protein interaction; curved DNA;
KW HIV replication; HIV integration; repair enzyme.
XX
OS Homo sapiens.
XX
PN FR2772046-A1.
XX
PD 11-JUN-1999.
XX
PF 09-DEC-1997; 97FR-00015536.
XX
PR 09-DEC-1997; 97FR-00015536.
(COMS) COMMISSARIAT ENERGIE ATOMIQUE.
Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
WPI; 1999-359999/31.
New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.
Claim 19; Page 37-38; 69pp; French.
The present sequence represents a human kin17 protein. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein are useful for regulating the interaction between proteins and curved DNA. The fragment can be used to block replication of HIV or its integration into the human genome or to target repair enzymes to curved DNA sites. Expression vectors for kin17 can be used for controlling cell proliferation
Sequence 393 AA;

Query Match 100.0%; Score 393; DB 2; Length 393; Best Local Similarity 100.0%; Pred. No. 0; Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1	MGKSDFLTPKAIANRIKSGKQLRWYQCMQKQCRDENGFKCHCMSESHORQLLASEN 60
DB 1	MGKSDFLTPKAIANRIKSGKQLRWYQCMQKQCRDENGFKCHCMSESHORQLLASEN 60
QY 61	PQPFMDYFSEEFNDLELLRRFGTKRVHNNIVYNEYISHREHIMNATQWETLTDFTK 120
DB 61	PQPFMDYFSEEFNDLELLRRFGTKRVHNNIVYNEYISHREHIMNATQWETLTDFTK 120
QY 121	WLGREGKCKVDTPKGYIYIDRDPETIRRQLEKKKKQDLDDDEKTAFTIEEQVRG 180
DB 121	WLGREGKCKVDTPKGYIYIDRDPETIRRQLEKKKKQDLDDDEKTAFTIEEQVRG 180
QY 181	LEGKEQEVPTFTLSRENDEEKVTNLSKGACSSGATSSKSTLGPALKITIGSSASVK 240
DB 181	LEGKEQEVPTFTLSRENDEEKVTNLSKGACSSGATSSKSTLGPALKITIGSSASVK 240
QY 241	RKSSQSSQTSQSKKKKKKSSALDEIMEIEEKKRTARTDYWLOPEIIVKIITKKLGEKYHK 300
DB 241	RKSSQSSQTSQSKKKKKKSSALDEIMEIEEKKRTARTDYWLOPEIIVKIITKKLGEKYHK 300
QY 301	KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAGKRILVNGYRGNEGTLSEI 360
DB 301	KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAGKRILVNGYRGNEGTLSEI 360
QY 361	NEKTFSAITVIETGPKGRVVEGIOYEDISKLA 393
DB 361	NEKTFSAITVIETGPKGRVVEGIOYEDISKLA 393
RESULT 2	
ADJ69659	
ID	ADJ69659 standard; protein; 393 AA.
XX	
AC	ADJ69659;
XX	
DT	06-MAY-2004 (first entry)
XX	
DE	Human heat mitochondrial protein as a therapeutic target SeqID1465.
XX	
KW	mitochondrial; human; screening assay; diabetes mellitus;
KW	Huntington's disease; osteoarthritis;
KW	Leber's hereditary optic neuropathy; LHON;
KW	mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW	myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW	neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;
KW	osteopathic; ophthalmological; cytostatic.
XX	
OS	Homo sapiens.
XX	
PN	WO2003087768-A2.
XX	
PD	23-OCT-2003.
XX	
PF	04-APR-2003; 2003WO-US010870.
XX	
PR	12-APR-2002; 2002US-0372843P.
PR	17-JUN-2002; 2002US-0389987P.
PR	20-SEP-2002; 2002US-0412418P.
XX	
PA	(MITO-) MITOKOR.
PA	(BUCK-) BUCK INST AGE RES.
XX	
PI	Chosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
PI	Warnock DE;
DR	WPI; 2003-845369/78.
XX	
PT	Identifying a mitochondrial target for drug screening assays and for

PT	treating diseases associated with altered mitochondrial function, comprises detecting a modified polypeptide in a sample and correlating with the disease.
PT	
XX	
PS	Claim 1; SEQ ID NO 1465; 180pp; English.
XX	
CC	This invention relates to novel mitochondrial targets that can be used for therapeutic intervention in treating a disease associated with altered mitochondrial function. Specifically, it refers to a method for identifying proteins of the human heart mitochondrial proteome that are useful for drug screening assays, as well as therapeutic targets. The present invention describes a method for identifying such proteins that can be used in the treatment of various diseases associated with altered mitochondrial function including diabetes mellitus, Huntington's disease, osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy ragged red fibre syndrome (MERRF) or cancer. Accordingly, these compositions have neuroprotective, nontropic, antidiabetic, anticonvulsant, antiarthritic, osteopathic, ophthalmological and cytostatic activities. This polypeptide sequence is a human heart mitochondrial protein of the invention.
CC	
XX	
SQ	Sequence 393 AA;
Query Match 100.0%; Score 393; DB 7; Length 393; Best Local Similarity 100.0%; Pred. No. 0; Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1	MGKSDFLTPKAIANRIKSGKQLRWYQCMQKQCRDENGFKCHCMSESHORQLLASEN 60
DB 1	MGKSDFLTPKAIANRIKSGKQLRWYQCMQKQCRDENGFKCHCMSESHORQLLASEN 60
QY 61	PQPFMDYFSEEFNDLELLRRFGTKRVHNNIVYNEYISHREHIMNATQWETLTDFTK 120
DB 61	PQPFMDYFSEEFNDLELLRRFGTKRVHNNIVYNEYISHREHIMNATQWETLTDFTK 120
QY 121	WLGREGKCKVDTPKGYIYIDRDPETIRRQLEKKKKQDLDDDEKTAFTIEEQVRG 180
DB 121	WLGREGKCKVDTPKGYIYIDRDPETIRRQLEKKKKQDLDDDEKTAFTIEEQVRG 180
QY 181	LEGKEQEVPTFTLSRENDEEKVTNLSKGACSSGATSSKSTLGPALKITIGSSASVK 240
DB 181	LEGKEQEVPTFTLSRENDEEKVTNLSKGACSSGATSSKSTLGPALKITIGSSASVK 240
QY 241	RKSSQSSQTSQSKKKKKKSSALDEIMEIEEKKRTARTDYWLOPEIIVKIITKKLGEKYHK 300
DB 241	RKSSQSSQTSQSKKKKKKSSALDEIMEIEEKKRTARTDYWLOPEIIVKIITKKLGEKYHK 300
QY 301	KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAGKRILVNGYRGNEGTLSEI 360
DB 301	KKAIVKEVIDKYTAVVKMIDSGDKLKDQTHLETVIPAGKRILVNGYRGNEGTLSEI 360
QY 361	NEKTFSAITVIETGPKGRVVEGIOYEDISKLA 393
DB 361	NEKTFSAITVIETGPKGRVVEGIOYEDISKLA 393
RESULT 3	
AAY23619	
ID	AAY23619 standard; protein; 391 AA.
XX	
AC	AAY23619;
XX	
DT	07-SEP-1999 (first entry)
XX	
DE	Murine kin17 protein.
XX	
KW	Mouse; kin17 protein; cell proliferation; fertility;
KW	hyperproliferative disease; protein interaction; curved DNA;
KW	HIV replication; HIV integration; repair enzyme.
OS	Mus sp.
XX	

PN FR272046-A1.
 XX 11-JUN-1999.
 XX 09-DEC-1997; 97FR-00015536.
 XX 09-DEC-1997; 97FR-00015536.
 XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 XX WPI; 1999-359999/31.
 XX New DNA coding for human kin17 protein - useful for controlling cell
 XX proliferation or fertility.
 XX Claim 19; Page 36-37; 69pp; French.
 XX The present sequence represents a murine kin17 protein. The mammalian
 CC kin17 protein is useful for preparing a medicament for controlling cell
 CC proliferation or for controlling fertility. The medicaments can also be
 CC used to treat hyperproliferative diseases. Fragments between amino acids
 CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
 CC kin17 protein are useful for regulating the interaction between proteins
 CC and curved DNA. The fragment can be used to block replication of HIV or
 CC its integration into the human genome or to target repair enzymes to
 CC curved DNA sites. Expression vectors for kin17 can be used for
 CC controlling cell proliferation
 XX Sequence 391 AA;
 XX Query Match 42.5%; Score 167; DB 2; Length 391;
 XX Best Local Similarity 100.0%; Pred. No. 6.4e-150;
 XX Matches 167; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 21 LOKLRWYCOMCOKCRDENGPKCHCMSESHORQLLASENPQQFMDYFSEFRNDFLELL 80
 Db 21 LOKLRWYCOMCOKCRDENGPKCHCMSESHORQLLASENPQQFMDYFSEFRNDFLELL 80
 Qy 81 RRRGTGKRVHNNIVYNEIYSHREIHMNATOWETLDTFTKWLGRGLCKVDETPKGYIQ 140
 Db 81 RRRGTGKRVHNNIVYNEIYSHREIHMNATOWETLDTFTKWLGRGLCKVDETPKGYIQ 140
 Qy 141 YIDRDPETIRQLLEKXKQDLDEKTAKFIEQVRRGLGKEQE 187
 Db 141 YIDRDPETIRQLLEKXKQDLDEKTAKFIEQVRRGLGKEQE 187
 RESULT 4
 AAY23618
 ID AAY23618 standard; protein; 293 AA.
 AC AAY23618;
 XX 07-SEP-1999 (first entry)
 XX A huamn truncated kin17 protein.
 XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 XX protein interaction; curved DNA; HIV replication; HIV integration;
 XX repair enzyme.
 XX Homo sapiens.
 XX FR272046-A1.
 XX 11-JUN-1999.
 XX 09-DEC-1997; 97FR-00015536.
 XX 09-DEC-1997; 97FR-00015536.
 XX PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 XX WPI; 1999-359999/31.
 XX New DNA coding for human kin17 protein - useful for controlling cell
 XX proliferation or fertility.
 XX Claim 15; Page 34-35; 69pp; French.
 XX The present sequence represents a truncated human kin17 protein with
 CC amino acids 129-228 deleted. The mammalian kin17 protein is useful for
 CC preparing a medicament for controlling cell proliferation or for
 CC controlling fertility. The medicaments can also be used to treat
 CC hyperproliferative diseases. Fragments between amino acids 55 and 235
 CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein
 CC are useful for regulating the interaction between proteins and curved
 CC DNA. The fragment can be used to block replication of HIV or its
 CC integration into the human genome or to target repair enzymes to curved
 CC DNA sites. Expression vectors for kin17 can be used for controlling cell
 CC proliferation
 XX Sequence 293 AA;
 XX Query Match 42.0%; Score 165; DB 2; Length 293;
 XX Best Local Similarity 100.0%; Pred. No. 4e-148;
 XX Matches 165; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 229 ALKTIGSSASVVRKESQSQSTQSKKKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVK 288
 Db 129 ALKTIGSSASVVRKESQSQSTQSKKKKKKSSALDEIMEIEEEKKRTARTDYWLQPEIIVK 188
 Qy 289 IITKLGKGYHKKKAIKVEIDKYTAVVWMDSGDKLKDQTHLETVPAPGKRILVNG 348
 Db 189 IITKLGKGYHKKKAIKVEIDKYTAVVWMDSGDKLKDQTHLETVPAPGKRILVNG 248
 Qy 349 GYRNGEGLTESINEKTFSAIVTETGPKGRRVEGQYEDISKLA 393
 Db 249 GYRNGEGLTESINEKTFSAIVTETGPKGRRVEGQYEDISKLA 293
 RESULT 5
 AAY23617
 ID AAY23617 standard; protein; 291 AA.
 AC AAY23617;
 XX 07-SEP-1999 (first entry)
 XX Mouse truncated kin17 protein.
 XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 XX protein interaction; curved DNA; HIV replication; HIV integration;
 XX repair enzyme.
 XX Mus sp.
 XX FR272046-A1.
 XX 11-JUN-1999.
 XX 09-DEC-1997; 97FR-00015536.
 XX 09-DEC-1997; 97FR-00015536.
 XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 XX WPI; 1999-359999/31.
 XX New DNA coding for human kin17 protein - useful for controlling cell
 XX proliferation or fertility.

```
PT proliferation or fertility.
XX
PS Claim 14; Page 33-34; 69pp; French.
XX
CC The present sequence represents a mouse kin17 protein with amino acids
CC 129-228 deleted. The mammalian kin17 protein is useful for preparing a
CC medicament for controlling cell proliferation or for controlling
CC fertility. The medicaments can also be used to treat hyperproliferative
CC diseases. Fragments between amino acids 55 and 235 (preferably between
CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
CC regulating the interaction between proteins and curved DNA. The fragment
CC can be used to block replication of HIV or its integration into the human
CC genome or to target repair enzymes to curved DNA sites. Expression
CC vectors for kin17 can be used for controlling cell proliferation
XX
SQ Sequence 291 AA;
Query Match 30.5%; Score 120; DB 2; Length 291;
Best Local Similarity 100.0%; Pred. No. 2.7e-105;
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 PRAIANRIKSGIKLQKLRWYCMQCKQCRDENGFKCHCMSESHQRLLLASENPQQPMDYF 68
DB 9 PRAIANRIKSGIKLQKLRWYCMQCKQCRDENGFKCHCMSESHQRLLLASENPQQPMDYF 68
QY 69 SEEFRNDFLELLRRRGTGKRVHNNIVYNEYISHREHHMNATQWETLDTFTKWLGREGLC 128
DB 69 SEEFRNDFLELLRRRGTGKRVHNNIVYNEYISHREHHMNATQWETLDTFTKWLGREGLC 128
RESULT 6
AAG74690
ID AAG74690 standard; protein; 84 AA.
XX
AC AAG74690;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:5454.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
WPI: 2001-235357/24.
XX
N-PSDB; AAH34095.
XX
Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers.
XX
Claim 11; Page 7064-7065; 9803pp; English.
XX
AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where the
XX proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene therapy
XX and vaccine production. N and P may be used in the prevention, diagnosis
XX and treatment of diseases associated with inappropriate P expression. For
XX example, N and P may be used to treat disorders associated with decreased
CC
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patient's own production of P. Additionally, N may be used
CC to produce the colon cancer-associated P, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 84 AA;
Query Match 18.6%; Score 73; DB 4; Length 84;
Best Local Similarity 100.0%; Pred. No. 4.8e-61;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 321 SGDKLKLDQTHLETVPAPGKRILVINGYRGNEGTLESINEKTSATIVETGPKGRR 380
DB 12 SGDKLKLDQTHLETVPAPGKRILVINGYRGNEGTLESINEKTSATIVETGPKGRR 71
QY 381 VEGIOVEDISKLA 393
DB 72 VEGIOVEDISKLA 84
RESULT 7
ADP22452
ID ADP22452 standard; protein; 390 AA.
XX
AC ADP22452;
XX
DT 12-AUG-2004 (first entry)
XX
DE Sea-squirt (Ciona intestinalis) zinc finger protein #13.
XX
KW sea-squirt; zinc finger protein; gene detection; drug development;
KW zinc finger protein-associated disease.
XX
OS Ciona intestinalis.
XX
PN JP2004057126-A.
XX
PD 26-FEB-2004.
XX
PF 31-JUL-2002; 2002JP-00222484.
XX
PR 31-JUL-2002; 2002JP-00222484.
XX
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX
WPI: 2004-208711/20.
XX
N-PSDB; ADP22451.
XX
Novel gene encoding zinc finger protein, useful as probe in gene
XX detecting instruments and in development of drug for treating zinc finger
XX protein associated diseases.
XX
Claim 1; SEQ ID NO 26; 972pp; Japanese.
XX
The invention comprises the amino acid and coding sequences of sea-squirt
XX (Ciona intestinalis) zinc finger proteins. The DNA and protein sequences
XX of the invention are useful in a gene detecting instrument. The DNA and
XX protein sequences of the invention are useful in the development of drugs
XX for the treatment of zinc finger protein-associated diseases. The present
XX amino acid sequence represents a sea-squirt zinc finger protein of the
XX invention.
XX
Sequence 390 AA;
Query Match 10.7%; Score 42; DB 8; Length 390;
Best Local Similarity 100.0%; Pred. No. 6.4e-31;
```


Matches	42;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	14	NRKSKGLQKLRWY	COMCQCRDENGFKCHCMSESHORQLL	55					
Db	16	NRKSKGLQKLRWY	COMCQCRDENGFKCHCMSESHORQLL	57					
RESULT 8									
AAG42576									
ID	AAG42576	standard; protein; 411 AA.							
XX	AC	AAG42576;							
XX	AC								
XX	DT	18-OCT-2000 (first entry)							
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 53114.							
XX	KW	Protein identification; signal transduction pathway; metabolic pathway;							
XX	KW	hybridisation assay; genetic mapping; gene expression control; promoter;							
XX	KW	termination sequence.							
XX	OS	Arabidopsis thaliana.							
XX	PN	EP1033405-A2.							
XX	PD	06-SEP-2000.							
XX	PF	25-FEB-2000; 2000EP-00301439.							
XX	PR	25-FEB-1999; 99US-0121825P.							
XX	PR	05-MAR-1999; 99US-0123180P.							
XX	PR	09-MAR-1999; 99US-0123548P.							
XX	PR	23-MAR-1999; 99US-0125788P.							
XX	PR	25-MAR-1999; 99US-0126264P.							
XX	PR	29-MAR-1999; 99US-0126785P.							
XX	PR	01-APR-1999; 99US-0127462P.							
XX	PR	06-APR-1999; 99US-0128234P.							
XX	PR	08-APR-1999; 99US-0128714P.							
XX	PR	16-APR-1999; 99US-0129845P.							
XX	PR	19-APR-1999; 99US-0130077P.							
XX	PR	21-APR-1999; 99US-0130449P.							
XX	PR	23-APR-1999; 99US-0130510P.							
XX	PR	23-APR-1999; 99US-0130891P.							
XX	PR	28-APR-1999; 99US-0131449P.							
XX	PR	30-APR-1999; 99US-0132048P.							
XX	PR	30-APR-1999; 99US-0132407P.							
XX	PR	04-MAY-1999; 99US-0132484P.							
XX	PR	05-MAY-1999; 99US-0132485P.							
XX	PR	06-MAY-1999; 99US-0132486P.							
XX	PR	06-MAY-1999; 99US-0132487P.							
XX	PR	07-MAY-1999; 99US-0132863P.							
XX	PR	11-MAY-1999; 99US-0134256P.							
XX	PR	14-MAY-1999; 99US-0134218P.							
XX	PR	14-MAY-1999; 99US-0134219P.							
XX	PR	14-MAY-1999; 99US-0134221P.							
XX	PR	14-MAY-1999; 99US-0134370P.							
XX	PR	18-MAY-1999; 99US-0134768P.							
XX	PR	19-MAY-1999; 99US-0134941P.							
XX	PR	20-MAY-1999; 99US-0135124P.							
XX	PR	21-MAY-1999; 99US-0135353P.							
XX	PR	24-MAY-1999; 99US-0135629P.							
XX	PR	25-MAY-1999; 99US-0136021P.							
XX	PR	27-MAY-1999; 99US-0136392P.							
XX	PR	28-MAY-1999; 99US-0136782P.							
XX	PR	01-JUN-1999; 99US-0137222P.							
XX	PR	03-JUN-1999; 99US-0137528P.							
XX	PR	04-JUN-1999; 99US-0137502P.							
XX	PR	07-JUN-1999; 99US-0137724P.							
XX	PR	08-JUN-1999; 99US-0138094P.							
XX	PR	10-JUN-1999; 99US-0138540P.							
XX	PR	10-JUN-1999; 99US-0138847P.							
XX	PR	14-JUN-1999; 99US-0139119P.							
XX	PR	16-JUN-1999; 99US-0139452P.							
XX	PR	16-JUN-1999; 99US-0139453P.							
XX	PR	17-JUN-1999; 99US-0139492P.							
XX	PR	18-JUN-1999; 99US-0139454P.							
XX	PR	18-JUN-1999; 99US-0139455P.							
XX	PR	18-JUN-1999; 99US-0139456P.							
XX	PR	18-JUN-1999; 99US-0139457P.							
XX	PR	18-JUN-1999; 99US-0139458P.							
XX	PR	18-JUN-1999; 99US-0139459P.							
XX	PR	18-JUN-1999; 99US-0139460P.							
XX	PR	18-JUN-1999; 99US-0139461P.							
XX	PR	18-JUN-1999; 99US-0139462P.							
XX	PR	18-JUN-1999; 99US-0139463P.							
XX	PR	18-JUN-1999; 99US-0139750P.							
XX	PR	18-JUN-1999; 99US-0139763P.							
XX	PR	21-JUN-1999; 99US-0139817P.							
XX	PR	22-JUN-1999; 99US-0139899P.							
XX	PR	23-JUN-1999; 99US-0140353P.							
XX	PR	23-JUN-1999; 99US-0140354P.							
XX	PR	24-JUN-1999; 99US-0140695P.							
XX	PR	28-JUN-1999; 99US-0140823P.							
XX	PR	29-JUN-1999; 99US-0140991P.							
XX	PR	30-JUN-1999; 99US-0141287P.							
XX	PR	01-JUL-1999; 99US-0141842P.							
XX	PR	02-JUL-1999; 99US-0142154P.							
XX	PR	02-JUL-1999; 99US-0142055P.							
XX	PR	06-JUL-1999; 99US-0142390P.							
XX	PR	08-JUL-1999; 99US-0142803P.							
XX	PR	09-JUL-1999; 99US-0142920P.							
XX	PR	12-JUL-1999; 99US-0142977P.							
XX	PR	13-JUL-1999; 99US-0143542P.							
XX	PR	14-JUL-1999; 99US-0143624P.							
XX	PR	15-JUL-1999; 99US-0144005P.							
XX	PR	16-JUL-1999; 99US-0144085P.							
XX	PR	16-JUL-1999; 99US-0144086P.							
XX	PR	19-JUL-1999; 99US-0144325P.							
XX	PR	19-JUL-1999; 99US-0144331P.							
XX	PR	19-JUL-1999; 99US-0144332P.							
XX	PR	19-JUL-1999; 99US-0144333P.							
XX	PR	19-JUL-1999; 99US-0144334P.							
XX	PR	19-JUL-1999; 99US-0144335P.							
XX	PR	20-JUL-1999; 99US-0144352P.							
XX	PR	20-JUL-1999; 99US-0144632P.							
XX	PR	20-JUL-1999; 99US-0144884P.							
XX	PR	21-JUL-1999; 99US-0144814P.							
XX	PR	21-JUL-1999; 99US-0145086P.							
XX	PR	21-JUL-1999; 99US-0145088P.							
XX	PR	22-JUL-1999; 99US-0145085P.							
XX	PR	22-JUL-1999; 99US-0145087P.							
XX	PR	22-JUL-1999; 99US-0145089P.							
XX	PR	22-JUL-1999; 99US-0145192P.							
XX	PR	23-JUL-1999; 99US-0145145P.							
XX	PR	23-JUL-1999; 99US-0145218P.							
XX	PR	23-JUL-1999; 99US-0145224P.							
XX	PR	26-JUL-1999; 99US-0145276P.							
XX	PR	27-JUL-1999; 99US-0145913P.							
XX	PR	27-JUL-1999; 99US-0145918P.							
XX	PR	27-JUL-1999; 99US-0145919P.							
XX	PR	28-JUL-1999; 99US-0145951P.							
XX	PR	02-AUG-1999; 99US-0146386P.							
XX	PR	02-AUG-1999; 99US-0146388P.							
XX	PR	02-AUG-1999; 99US-0146389P.							
XX	PR	03-AUG-1999; 99US-0147038P.							
XX	PR	04-AUG-1999; 99US-0147204P.							
XX	PR	04-AUG-1999; 99US-0147302P.							
XX	PR	05-AUG-1999; 99US-0147192P.							
XX	PR	05-AUG-1999; 99US-0147260P.							
XX	PR	06-AUG-1999; 99US-0147303P.							
XX	PR	06-AUG-1999; 99US-0147416P.							
XX	PR	09-AUG-1999; 99US-0147493P.							
XX	PR	09-AUG-1999; 99US-0147935P.							
XX	PR	10-AUG-1999; 99US-0148171P.							
XX	PR	11-AUG-1999; 99US-0148319P.							
XX	PR	12-AUG-1999; 99US-0148341P.							

```
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 22-OCT-1999; 99US-0161404P.
PR 23-OCT-1999; 99US-0161405P.
PR 23-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      8.9%; Score 35; DB 3; Length 411;
Best Local Similarity 100.0%; Pred. No. 3.1e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQRQ 53
   |||||
DB 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQRQ 53

RESULT 10
```

RESULT 9

AAB03064

ID AAB03064 standard; protein; 423 AA.

XX AAB03064;

XX 27-SEP-2000 (first entry)

XX Maize KIN17 orthologue, ZmKINH-2.

XX ZmKINH-2; KIN17 orthologue; maize; zinc finger protein; RecA homologue; nuclear localisation; nonhomologous recombination;

KW illegitimate recombination; double stranded DNA binding; curved DNA;

KW homologous gene targeting; transgenic plant.

XX Zea mays.

XX WO200024900-A1.

XX 04-MAY-2000.

XX 06-OCT-1999; 99WO-US023280.

XX 27-OCT-1998; 98US-0105802P.

XX (PION-) PIONEER HI-BRED INT INC.

XX Mahajan PB;

XX WPI; 2000-350741/30.

XX N-PSDB; AAA52590.

XX Nucleic acids encoding maize KIN17 orthologue proteins useful for preventing illegitimate recombination in cells.

XX Claim 11; Page 64-66; 84pp; English.

XX This sequence represents the maize KIN17 orthologue ZmKINH-2. The invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic acids encoding them (AAA52589-A52591), and expression vectors, transgenic plants and plant seeds comprising nucleotides encoding maize KIN17 orthologues. KIN17 has, until now, been found only in animal (avian, rodent and human) cells, this invention being the first report describing the presence of KIN17 in plants. Murine KIN17 was found to have a significant homology to Escherichia coli RecA protein, and contains a zinc finger motif and a nuclear localisation signal. KIN17 binds double-stranded DNA, preferentially binding to curved DNA, and forms intranuclear foci on overexpression in mammalian cells. It is also induced on exposure to gamma or ultraviolet radiation. These findings indicate that KIN17 plays a role in non-homologous (illegitimate) recombination, which occurs at higher rates among higher eukaryotes, particularly plants. Illegitimate recombination in plants is a major impediment to the generation of transgenic crops such as maize. Maize KIN17 orthologue nucleic acid sequences may be used to generate transgenic plants. The transgenic plants generated can be monocots or dicots and are particularly maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense sequences may be used to reduce KIN17 levels in embryogenic callus or embryo cells, thereby reducing the amount of non-homologous recombination and enhancing homologous gene targeting

XX Sequence 423 AA;

Query Match 8.9%; Score 35; DB 3; Length 423;

Best Local Similarity 100.0%; Pred. No. 3.2e-24;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQRQ 53

DB 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQRQ 53

AAB03065
 ID AAB03065 standard; protein; 423 AA.
 AC AAB03065;
 XX
 DT 27-SEP-2000 (first entry)
 XX
 DE Maize KIN17 orthologue, ZmKINH-3.
 XX
 KW ZmKINH-3; KIN17 orthologue; maize; zinc finger protein; RecA homologue;
 KW nuclear localisation; nonhomologous recombination;
 KW illegitimate recombination; double stranded DNA binding; curved DNA;
 KW homologous gene targeting; transgenic plant.
 OS
 XX Zea mays.
 XX
 XX WO200024900-A1.
 XX
 XX 04-MAY-2000.
 XX
 XX 06-OCT-1999; 99WO-US023280.
 PF
 XX 27-OCT-1998; 98US-0105802P.
 PR
 XX (PION-) PIONEER HI-BRED INT INC.
 XX
 XX Mahajan PB;
 PI
 XX WPI; 2000-350741/30.
 DR
 XX N-PSDB; AAA52591.
 DR
 XX
 XX Nucleic acids encoding maize KIN17 orthologue proteins useful for
 PT preventing illegitimate recombination in cells.
 PT
 XX
 XX Claim 11; Page 64-66; 84pp; English.
 XX
 XX This sequence represents the maize KIN17 orthologue ZmKINH-3. The
 CC invention relates to maize KIN17 orthologues (AAB03063-B03065), nucleic
 CC acids encoding them (AAA52589-A52591), and expression vectors, transgenic
 CC plants and plant seeds comprising nucleotides encoding maize KIN17
 CC orthologues. KIN17 has, until now, been found only in animal (avian,
 CC rodent and human) cells, this invention being the first report describing
 CC the presence of KIN17 in plants. Murine KIN17 was found to have
 CC significant homology to Escherichia coli RecA protein, and contains a
 CC zinc finger motif and a nuclear localisation signal. KIN17 binds double-
 CC stranded DNA, preferentially binding to curved DNA, and forms
 CC intranuclear foci on overexpression in mammalian cells. It is also
 CC induced on exposure to gamma or ultraviolet radiation. These findings
 CC indicate that KIN17 plays a role in non-homologous (illegitimate)
 CC recombination, which occurs at higher rates among higher eukaryotes,
 CC particularly plants. Illegitimate recombination in plants is a major
 CC impediment to the generation of transgenic crops such as maize. Maize
 CC KIN17 orthologue nucleic acid sequences may be used to generate
 CC transgenic plants. The transgenic plants generated can be monocots or
 CC dicots and are particularly maize, soybean, sunflower, sorghum, canola,
 CC wheat, alfalfa, cotton, rice, barley and millet. In particular, antisense
 CC sequences may be used to reduce KIN17 levels in embryogenic callus or
 CC embryo cells, thereby reducing the amount of non-homologous recombination
 CC and enhancing homologous gene targeting
 XX
 XX Sequence 423 AA;
 SQ
 Query Match 8.9%; Score 35; DB 3; Length 423;
 Best Local Similarity 100.0%; Pred. No. 3.2e-24;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 19 KGLQKLRYCQMCQRCQRCDENGFKCHCMSESHQ 53
 DB 19 KGLQKLRYCQMCQRCQRCDENGFKCHCMSESHQ 53
 RESULT 11
 ADT58422

ID ADT58422 standard; protein; 423 AA.
 AC ADT58422;
 XX
 DT 13-JAN-2005 (first entry)
 XX
 DE Plant polypeptide, SEQ ID 8499.
 XX
 KW plant; transgenic; cold tolerance; growth rate; drought tolerance;
 KW disease resistance; galactomannan production; plant growth regulator;
 KW heat tolerance; herbicide tolerance; lignin production;
 KW extreme osmotic condition tolerance; pathogens resistance;
 KW pest resistance; yield improvement; seed oil yield; seed protein yield.
 OS
 XX Viridiplantae.
 XX
 XX US2004216190-A1.
 XX
 XX 28-OCT-2004.
 XX
 XX 18-DEC-2003; 2003US-00739930.
 PF
 XX 28-APR-2003; 2003US-00424599.
 PR
 XX 28-APR-2003; 2003US-00425115.
 PR
 XX (KOVA/) KOVALIC D K.
 PA
 XX Kovalic DK;
 PI
 XX WPI; 2004-757369/74.
 DR
 XX
 XX New recombinant DNA constructs useful in the field of biochemistry and
 PT genetics, and in particular for producing transgenic plants with improved
 PT biological characteristics.
 PT
 XX
 XX Claim 2; SEQ ID NO 8499; 14pp; English.
 XX
 XX The invention relates a recombinant DNA construct comprising a
 CC polynucleotide having any of 5544 nucleotide sequences (CDNAS SEQ ID NO:
 CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences
 CC (SEQ ID NO: 5545-11088). The CDNAS and proteins are from corn, soybean,
 CC Arabidopsis, wheat and rape but the specification does not indicate which
 CC sequences is derived from which organism. Also included is a method of
 CC producing a plant having an improved property, comprising transforming a
 CC plant with a recombinant DNA construct comprising a promoter region
 CC functional in a plant cell operably joined to a polynucleotide encoding a
 CC polypeptide associated with the property, and growing the transformed
 CC plant. The property is selected from improving plant cold tolerance, for
 CC manipulating growth rate in plant cells by modification of the cell cycle
 CC pathway, for improving plant drought tolerance, for providing increased
 CC resistance to plant disease, for galactomannan production, for production
 CC of plant growth regulators, for improving plant heat tolerance, for
 CC improving plant tolerance to herbicides, for increasing the rate of
 CC homologous recombination in plants, for lignin production, for improving
 CC plant tolerance to extreme osmotic conditions, for improving plant
 CC tolerance to pathogens or pests, for yield improvement by modification of
 CC photosynthesis, for modifying seed oil yield and/or content, for
 CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake
 CC and for yield improvement by providing improved plant growth and
 CC development under at least one stress condition. The polynucleotide may
 CC also encode a plant transcription factor. The methods and compositions of
 CC the present invention are useful in the field of biochemistry and
 CC genetics, in particular for producing transgenic plants with improved
 CC biological characteristics such as increased yield, improved nitrogen
 CC flow, increasing plant tolerance to cold or heat, improving plant
 CC tolerance to extreme osmotic and drought conditions, and improving plant
 CC tolerance to plant pests or pathogens. They can also be used in physical
 CC arrays of molecules, plant breeding markers, computer-based storage and
 CC analysis systems. The present sequence is one of the 5544 plant protein
 CC sequences of the invention. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from USPTO at

CC or by providing improved plant growth and development under at least one
 CC stress condition or for modifying seed oil or protein yield and/or
 CC content. This is the amino acid sequence of a plant full length insert
 CC polypeptide that can be used in the recombinant DNA construct of the
 CC invention.

XX SQ Sequence 437 AA;

Query Match 8.9%; Score 35; DB 8; Length 437;
 Best Local Similarity 100.0%; Pred. No. 3.3e-24;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQ 53
 |||||
 Db 32 KGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQ 66

RESULT 14

AD78817
 ID ADX78817 standard; protein; 437 AA.

XX AC ADX78817;

XX DT 21-APR-2005 (first entry)

XX DE Plant full length insert polypeptide seqid 48183.

XX KW plant protectant; plant growth regulant; gene therapy; plant;
 KW recombinant DNA construct; physical array; plant breeding marker;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
 KW extreme osmotic condition; pathogen tolerance; pest tolerance;
 KW growth rate; cell cycle pathway; disease resistance;
 KW galactomannan production; lignin production; plant growth regulator;
 KW yield; plant growth; plant development; seed oil; protein yield;
 KW protein content.

XX OS Unidentified.

XX PN US2004034888-A1.

XX PD 19-FEB-2004.

XX PF 28-APR-2003; 2003US-00425114.

XX PR 06-MAY-1999; 99US-00304517.

XX PR 05-NOV-2001; 2001US-00985678.

XX PA (LIU/J) LIU J.

XX PA (ZHOU/J) ZHOU Y.

XX PA (KOVA/) KOVALIC D K.

XX PA (SCRE/) SCREEN S E.

XX PA (TAB/) TABASKA J E.

XX PA (CAOV/) CAO Y.

XX PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;

XX DR WPI; 2004-180133/17.

XX PT New recombinant DNA construct, useful for improving plant tolerance to
 PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
 PT pests, for conferring increased resistance to plant disease, or for
 PT improving yield.

XX PS Claim 1; SEQ ID NO 48183; 15pp; English.

XX CC The invention describes a recombinant DNA construct comprising a
 CC polynucleotide consisting of a sequence encoding an amino acid sequence
 CC available in electronic form from the US patent office at
 CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
 CC of the invention are also useful in physical arrays of molecules and as
 CC plant breeding markers. The recombinant DNA construct is useful for
 CC improving plant tolerance to cold, heat, drought, herbicides, extreme
 CC osmotic conditions, pathogens or pests, for manipulating growth rate in

CC plant cells by modification of the cell cycle pathway, for conferring
 CC increased resistance to plant disease, for producing galactomannan,
 CC lignin or plant growth regulators, for increasing the rate of homologous
 CC recombination in plants, for improving yield by modification of
 CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
 CC or by providing improved plant growth and development under at least one
 CC stress condition or for modifying seed oil or protein yield and/or
 CC content. This is the amino acid sequence of a plant full length insert
 CC polypeptide that can be used in the recombinant DNA construct of the
 CC invention.

XX SQ Sequence 437 AA;

Query Match 8.9%; Score 35; DB 8; Length 437;
 Best Local Similarity 100.0%; Pred. No. 3.3e-24;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQ 53
 |||||
 Db 32 KGLQKLRWYCMQCKQCRDENGFKCHCMSESHORQ 66

RESULT 15

AAG42577

ID AAG42577 standard; protein; 382 AA.

XX AC AAG42577;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 53115.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridization assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

XX PR 03-MAR-1999; 99US-0123380P.

XX PR 09-MAR-1999; 99US-0123548P.

XX PR 23-MAR-1999; 99US-0125788P.

XX PR 25-MAR-1999; 99US-0126264P.

XX PR 29-MAR-1999; 99US-0126785P.

XX PR 01-APR-1999; 99US-0127462P.

XX PR 06-APR-1999; 99US-0128234P.

XX PR 08-APR-1999; 99US-0128714P.

XX PR 16-APR-1999; 99US-0129845P.

XX PR 19-APR-1999; 99US-0130077P.

XX PR 21-APR-1999; 99US-0130449P.

XX PR 23-APR-1999; 99US-0130510P.

XX PR 28-APR-1999; 99US-0130891P.

XX PR 28-APR-1999; 99US-0131449P.

XX PR 30-APR-1999; 99US-0132048P.

XX PR 30-APR-1999; 99US-0132407P.

XX PR 04-MAY-1999; 99US-0132484P.

XX PR 05-MAY-1999; 99US-0132485P.

XX PR 06-MAY-1999; 99US-0132486P.

XX PR 06-MAY-1999; 99US-0132487P.

XX PR 07-MAY-1999; 99US-0132863P.

XX PR 11-MAY-1999; 99US-0134256P.

XX PR 14-MAY-1999; 99US-0134218P.

XX PR 14-MAY-1999; 99US-0134219P.

XX PR 14-MAY-1999; 99US-0134221P.

XX PR 14-MAY-1999; 99US-0134370P.

XX PR 18-MAY-1999; 99US-0134768P.

XX PR 19-MAY-1999; 99US-0134941P.

PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145193P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.

PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148315P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 06-OCT-1999; 99US-0157533P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.

this Page Blank (uspic)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:34:39 ; Search time 154.894 Seconds
(without alignments)
1790.078 Million cell updates/sec

Title: US-09-555-529-26

Perfect score: 393

Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPIKGRVBEQIYEDISKLA 393

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2166443 seqs, 705528306 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	393	100.0	393	2	O60870_HUMAN	O60870 homo sapien
2	179	45.5	195	2	Q9CV58_MOUSE	Q9CV58 mus musculus
3	179	45.5	391	2	Q8K339_MOUSE	Q8K339 mus musculus
4	65	16.5	389	2	O6GL97_XENTR	O6GL97 xenopus tro
5	50	12.7	387	2	O68F56_XENLA	O68F56 xenopus lae
6	48	12.2	427	2	O4S5G2_TETNG	O4S5G2 tetraodon n
7	43	10.9	383	2	O5RH25_BRARE	O5RH25 brachydanio
8	41	10.4	269	2	O4TDV9_TETNG	O4TDV9 tetraodon n
9	35	8.9	402	2	Q9AY87_ORYSA	Q9AY87 oryza sativ
10	35	8.9	411	2	Q9ZVU5_ARATH	Q9ZVU5 arabidopsis
11	35	8.9	430	2	O75LU5_ORYSA	O75LU5 oryza sativ
12	29	7.4	88	2	O4TDW0_TETNG	O4TDW0 tetraodon n
13	15	3.8	178	2	O6X190_DROPHI	O6X190 drosophila
14	15	3.8	244	2	O76926_DROME	O76926 drosophila
15	15	3.8	390	2	O8SXR2_DROME	O8SXR2 drosophila
16	15	3.8	390	2	O9VPH4_DROME	O9VPH4 drosophila
17	14	3.6	263	2	O5CFM3_CRYOP	O5CFM3 cryptospori
18	14	3.6	265	2	O5CYD1_CRYOP	O5CYD1 cryptospori
19	13	3.3	387	2	O7PGN1_ANOGA	O7PGN1 anopheles g
20	12	3.1	121	2	O4XND8_PLACH	O4XND8 plasmodium
21	12	3.1	239	2	O6CXJ7_KLUFA	O6CXJ7 kluyveromyc
22	12	3.1	378	2	O4N9K7_THEPA	O4N9K7 theileria p
23	12	3.1	412	2	O4UG63_THEAN	O4UG63 theileria a
24	12	3.1	441	2	O4XDV6_PLACH	O4XDV6 plasmodium
25	12	3.1	442	2	O8IKG2_PLAF7	O8IKG2 plasmodium
26	12	3.1	445	2	O55DI6_DICDI	O55DI6 dictyosteli
27	12	3.1	445	2	O4Z7D7_PLABE	O4Z7D7 plasmodium
28	12	3.1	457	2	O7RD49_PLAYO	O7RD49 plasmodium
29	11	2.8	238	2	O51F62_ENTHI	O51F62 entamoeba h
30	10	2.5	397	2	O60L39_CAEBR	O60L39 caenorhabdi
31	10	2.5	404	2	Q9XWF2_CAEEL	Q9XWF2 caenorhabdi

RESULT 1						
O60870_HUMAN						
ID	O60870_HUMAN	PRELIMINARY;	PRT;	393	AA.	
AC	O60870;					
DT	01-AUG-1998	(TREMblrel. 07, Created)				
DT	01-JAN-1999	(TREMblrel. 09, Last sequence update)				
DT	13-SEP-2005	(TREMblrel. 31, Last annotation update)				
DE	Kin17 protein (HsKin17 protein) (KIN, antigenic determinant of recA protein homolog) (Mouse).					
DE	Name=KIN; Synonyms=Kin17; ORFNames=RP11-264C14.1-001;					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;					
OC	Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	NUCLEOTIDE SEQUENCE.					
RC	TISSUE=Testis;					
RX	MEDLINE=20419742; PubMed=10964102; DOI=10.1093/carcin/21.9.1701;					
RA	Kannouche P., Mauffrey P., Pinon-Lataillade G., Mattei M.G.,					
RA	Sarasin A., Daya-Grosjean L., Angulo J.F.;					
RT	"Molecular cloning and characterization of the human KIN17 cDNA encoding a component of the UVC response that is conserved among metazoans."					
RT	Carcinogenesis 21:1701-1710(2000).					
RL						
RN						
RP	NUCLEOTIDE SEQUENCE.					
RC	TISSUE=Pancreas;					
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;					
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,					
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,					
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,					
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haeh F.,					
RA	Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,					
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,					
RA	Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,					
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,					
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,					
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,					
RA	Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,					
RA	Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,					
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,					
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,					
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,					
RA	Butterfield J.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,					
RA	Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;					
RT	"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."					
RT	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).					
RL						
RN						
RP	NUCLEOTIDE SEQUENCE.					
RC	TISSUE=Pancreas;					
RA	Strausberg R.;					
RL	Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.					

Q9Y7X9 schizosacch
Q55M94 cryptococcu
Q5K9H7 cryptococcu
Q4W6G0 aspergillus
Q5AQ80 aspergillus
Q54XN1 dictyosteli
Q8PYN9 methanosarc
Q4NX48 anaeromyxob
Q6N18 xenopus lae
Q5GWA7 idiomarina
Q8XVB3 ralatonia s
P77992 thermococcu
Q9UY24 pyrococcus
Q8U438 pyrococcus

```

RN
RP NUCLEOTIDE SEQUENCE.
RA Lovell J.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ005273; CA00462.1; -; mRNA.
DR EMBL; BC017309; AAH17309.1; -; mRNA.
DR EMBL; AL158044; CAI12959.1; -; Genomic DNA.
DR Ensembl; ENSG00000151657; Homo sapiens.
DR HGNC; HGNC:6327; KIN.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003677; F:DNA binding; TAS.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF00467; KOW; 1.
DR SMART; SM00355; ZnF_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
SQ NUCLEAR PROTEIN.
KW SEQUENCE 393 AA; 45374 MW; 515A89B4C8A4C007 CRC64;

Query Match 100.0%; Score 393; DB 2; Length 393;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 393; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGKSDFLTPKAIANRIKSKGLQKLRWYQCMQKQCRDENGFKCHCMSESHQROLILASEN 60
DB 1 MGKSDFLTPKAIANRIKSKGLQKLRWYQCMQKQCRDENGFKCHCMSESHQROLILASEN 60

QY 61 PQQFMDYFSEPRNDFLELLRRRFGTKRVHNNIVYNEYISHREHHMNTQWETLTDFTK 120
DB 61 PQQFMDYFSEPRNDFLELLRRRFGTKRVHNNIVYNEYISHREHHMNTQWETLTDFTK 120

QY 121 WLGRGLCKVDTPKQWYQYIDRDPTTRQLELEKKKKQDLDDDEKTAFTFEEQVRG 180
DB 121 WLGRGLCKVDTPKQWYQYIDRDPTTRQLELEKKKKQDLDDDEKTAFTFEEQVRG 180

QY 181 LEGKEQVPTFTLSRENDEKVTNLSKGCSSGATSKSSTLGPALKITGSSASVK 240
DB 181 LEGKEQVPTFTLSRENDEKVTNLSKGCSSGATSKSSTLGPALKITGSSASVK 240

QY 241 RKSSQSSQSKKKKKKSSALDEIMBIEBKRTARTDYWLQPEIIVKIITKKLGEKYHK 300
DB 241 RKSSQSSQSKKKKKKSSALDEIMBIEBKRTARTDYWLQPEIIVKIITKKLGEKYHK 300

QY 301 KKAIVKEVDKYTAVVKMIDSGDKLQDQTHLETVPAPGKRILVNLGGYRGNEGTLESI 360
DB 301 KKAIVKEVDKYTAVVKMIDSGDKLQDQTHLETVPAPGKRILVNLGGYRGNEGTLESI 360

QY 361 NEKTFSAIVIEGTPLKGRVEGIQVEDISKLA 393
DB 361 NEKTFSAIVIEGTPLKGRVEGIQVEDISKLA 393

RESULT 2
Q9CV58_MOUSE
ID Q9CV58_MOUSE PRELIMINARY; PRT; 195 AA.
AC Q9CV58;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male tongue cDNA, RIKEN full-length enriched
DE library, clone:2310020017 product:antigenic determinant of rec-A
DE protein, full insert sequence. (fragment).
GN Names=Kin;
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK009429; BAB26281.2; -; mRNA.

"High-efficiency full-length cDNA cloning.";
Meth. Enzymol. 303:19-44(1999).
[2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RL MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Konno H., Yoshino M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RL The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RL MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RL MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RL ADACHI J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Haneigaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai K., Sakai T.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK009429; BAB26281.2; -; mRNA.
```

RP NUCLEOTIDE SEQUENCE:
RX MEDLINE=92020193; PubMed=19237963

```

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC074606; AAH74606.1; -; mRNA.
DR GO; GO:0030529; C:ribonucleoprotein complex; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN 1.
DR GO; GO:0003270; F:zinc ion binding; IEA.
SQ SEQUENCE 389 AA; 44987 MW; 793E70C997451FEA CRC64;

Query Match 16.5%; Score 65; DB 2; Length 389;
Best Local Similarity 100.0%; Pred. No. 1e-56;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 DFLTPKAIANRIKSGLOKLRWYCMQCQKCRDENGFKCHMSHSHORQLLLASENPQOF 64
DB 5 DFLTPKAIANRIKSGLOKLRWYCMQCQKCRDENGFKCHMSHSHORQLLLASENPQOF 64

QY 65 MDYFS 69
DB 65 MDYFS 69

RESULT 5
Q68F56_XENLA
ID Q68F56_XENLA PRELIMINARY; PRT; 387 AA.
AC Q68F56;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE WGC81626 protein.
GN Names=WGC81626;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heiton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human

```

```

RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC079988; AAH79988.1; -; mRNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0030529; C:ribonucleoprotein complex; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR005824; KOW.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00467; KOW; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
SQ SEQUENCE 387 AA; 44491 MW; 4CCB56B047F3779 CRC64;

Query Match 12.7%; Score 50; DB 2; Length 387;
Best Local Similarity 100.0%; Pred. No. 1.8e-41;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 NRIKSGLOKLRWYCMQCQKCRDENGFKCHMSHSHORQLLLASENPQ 63
DB 14 NRIKSGLOKLRWYCMQCQKCRDENGFKCHMSHSHORQLLLASENPQ 63

RESULT 6
Q4S5G2_TETNG
ID Q4S5G2_TETNG PRELIMINARY; PRT; 427 AA.
AC Q4S5G2;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome 19 SCAP14731, whole genome shotgun sequence.
GN ORFNames=GSTENG00023750001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segreus B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
RA Parra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier P., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957 (2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAEE0104731; CAG04120.1; -; Genomic DNA.
SQ SEQUENCE 427 AA; 48007 MW; ADF91299BEC23B CRC64;

Query Match 12.2%; Score 48; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 2.1e-39;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
QY 14 NRKSGKGLRWYCMCKQCRDENGFKCHMSHSHORQLLASNP 61
Db 14 NRKSGKGLRWYCMCKQCRDENGFKCHMSHSHORQLLASNP 61

RESULT 7
Q5RH25_BRARE
ID Q5RH25_BRARE PRELIMINARY; PRT; 383 AA.
AC Q5RH25;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DE Novel protein similar to vertebrate KIN, antigenic determinant of reca
DE protein homolog (Mouse) (KIN).
GN Name=OTTDARP0000008158; ORFNames=CH211-235A22.3-001;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN 1;
RP NUCLEOTIDE SEQUENCE.
RA Hunter G.;
RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX640594; CAIL1855.1; -; Genomic_DNA.
DR InterPro; IPR005824; KOW.
DR Pfam; PF00467; KOW; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKOWN 1.
SQ SEQUENCE 383 AA; 44337 MW; 1F0AA34FB62176D1 CRC64;

Query Match 10.9%; Score 43; DB 2; Length 383;
Best Local Similarity 100.0%; Pred. No. 2.3e-34;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLWYCMCKQCRDENGFKCHMSHSHORQLLASNP 61
Db 19 KGLQKLWYCMCKQCRDENGFKCHMSHSHORQLLASNP 61

RESULT 8
Q4TDV9_TETNG
ID Q4TDV9_TETNG PRELIMINARY; PRT; 269 AA.
AC Q4TDV9;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF6005, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG0002601001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN 1;
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.M., Brunet P., Petit J.L., Stange-Thomann N.,
RA Maucell E., Bouteau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anhouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Bieumont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chappier C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Winkler P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
the early vertebrate proto-karyotype."
```

```
RL Nature 431:946-957 (2004) .
RN 1;
RP NUCLEOTIDE SEQUENCE.
RG Genoscope, Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAE01006005; CAF88923.1; -; Genomic_DNA.
FT NON_TER 1
FT NON_TER 269
SQ SEQUENCE 269 AA; 31736 MW; E6883839A2CD17AF CRC64;

Query Match 10.4%; Score 41; DB 2; Length 269;
Best Local Similarity 100.0%; Pred. No. 1.8e-32;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 WETLTDTFKWLGREGICKVDETPKGWYQIYIDRDPETIRQ 152
Db 58 WETLTDTFKWLGREGICKVDETPKGWYQIYIDRDPETIRQ 98

RESULT 9
Q9AY87_ORYSA
ID Q9AY87_ORYSA PRELIMINARY; PRT; 402 AA.
AC Q9AY87;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein OSUNBa0004B24.17.
GN Name=OSUNBa0004B24.17;
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartodeae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN 1;
RP NUCLEOTIDE SEQUENCE.
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hailao J.,
RA Zismann V., Pal G., Bowman C.L., Fujii C.Y., Vanaken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN 2;
RP NUCLEOTIDE SEQUENCE.
RA Buell R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC084319; AAG59654.1; -; Genomic_DNA.
DR Gramene; Q9AY87; -;
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKOWN 1.
KW Hypothetical protein.
SQ SEQUENCE 402 AA; 45962 MW; A1F4E38ECC60D13E CRC64;

Query Match 8.9%; Score 35; DB 2; Length 402;
Best Local Similarity 100.0%; Pred. No. 3.2e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLWYCMCKQCRDENGFKCHMSHSHORQ 53
Db 19 KGLQKLWYCMCKQCRDENGFKCHMSHSHORQ 53

RESULT 10
Q9ZVU5_ARATH
ID Q9ZVU5_ARATH PRELIMINARY; PRT; 411 AA.
AC Q9ZVU5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
```

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 DE T5A14.13 protein (Hypothetical protein At1g55460).
 GN Names=T5A14.13; Synonyms=At1g55460;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
 RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
 RA Gonzalez A., Kremenetskaia I., Kim C., Lenz C., Li J., Liu S.,
 RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
 RA Walker M., Yu G., Ecker J.R., Theologis A., Davis R.W.;
 RN Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Etgu P., Lee J.M.,
 RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
 RN Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP NUCLEOTIDE SEQUENCE.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Toriumi M., Yu G., Bowser L.,
 RA Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Lam B., Lin J.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M.,
 RA Seki M., Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
 RA Theologis A.;
 RN Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AC005223; AAD10649.1; -; Genomic_DNA.
 DR EMBL; AY051011; AAK93688.1; -; mRNA.
 DR EMBL; AF360132; AAK25842.1; -; mRNA.
 DR PIR; H9596; H96596.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR006646; KOW-sub.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00467; KOW; 1.
 DR SMART; SM00739; KOW; 1.
 DR SMART; SM00355; Znf_C2H2; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 411 AA; 47288 MW; 9DA6F8648002065D CRC64;
 Query Match 8.9%; Score 35; DB 2; Length 411;
 Best Local Similarity 100.0%; Pred. No. 3.3e-26;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQRQ 53
 Db 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQRQ 53
 RESULT 11
 Q75LU5_ORYSA
 ID Q75LU5_ORYSA PRELIMINARY; PRT; 430 AA.
 AC Q75LU5;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DE Hypothetical protein OSUNB0015121.3.
 GN Names=OSUNB0015121.3;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.

OX NCBI_TaxID=39947;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
 RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,
 RA Fadrosch D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,
 RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feidblyum T.V.,
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
 RA White O., Salzberg S.L., Fraser C.W.;
 RN Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell R.;
 RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC091302; AAR00634.1; -; Genomic_DNA.
 DR Gramene; Q75LU5; -;
 DR InterPro; IPR005824; KOW.
 DR InterPro; IPR007087; Znf_C2H2.
 DR Pfam; PF00467; KOW; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 430 AA; 49067 MW; 6151D748ED2F4690 CRC64;
 Query Match 8.9%; Score 35; DB 2; Length 430;
 Best Local Similarity 100.0%; Pred. No. 3.4e-26;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQRQ 53
 Db 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHQRQ 53
 RESULT 12
 Q4TDW0_TETNG
 ID Q4TDW0_TETNG PRELIMINARY; PRT; 88 AA.
 AC Q4TDW0;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Chromosome undetermined SCAF6004, whole genome shotgun sequence.
 DE (Fragment).
 DE ORFNames=GSTENG0002600001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fiecher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Blement C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
 RA Farra G., Lardier G., Chapelle C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
 RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype.";
 RN Nature 431:946-957(2004).
 RL [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL; CAAE01006004; CAF88922.1; -; Genomic_DNA.
 DR

Search completed: November 23, 2005, 16:49:38
Job time : 155.894 secs

This Page Blank (uspic)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: November 23, 2005, 16:35:49 ; Search time 33.0842 Seconds
(without alignments)
982.087 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 393
Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPKGRVRGQIYEDISKLA 393

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 572060 seqs, 82675679 residues

Word size : 0

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:..
1: /cgn2_6/ptodata/1/iaa/5 COMB.pdp.*
2: /cgn2_6/ptodata/1/iaa/6 COMB.pdp.*
3: /cgn2_6/ptodata/1/iaa/H COMB.pdp.*
4: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pdp.*
5: /cgn2_6/ptodata/1/iaa/RE COMB.pdp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	2.3	401	2	US-09-270-767-44680
2	8	2.0	55	2	US-09-513-999C-6360
3	8	2.0	61	2	US-09-149-476-615
4	8	2.0	136	2	US-09-270-767-32008
5	8	2.0	176	1	US-08-809-267-10
6	8	2.0	176	2	US-10-135-807-8
7	8	2.0	176	4	PCT-US95-13662A-10
8	8	2.0	178	2	US-10-135-807-6
9	8	2.0	178	2	US-10-135-807-7
10	8	2.0	984	2	US-09-171-461-14
11	8	2.0	984	2	US-09-970-711-14
12	7	1.8	61	2	US-09-248-796A-22505
13	7	1.8	63	2	US-08-971-089-2
14	7	1.8	63	2	US-09-513-999C-5320
15	7	1.8	63	2	US-10-117-604A-2
16	7	1.8	64	2	US-09-248-796A-26971
17	7	1.8	67	2	US-09-248-796A-24256
18	7	1.8	68	2	US-09-270-767-34723
19	7	1.8	68	2	US-09-270-767-49940
20	7	1.8	69	2	US-09-107-433-4761
21	7	1.8	76	2	US-09-248-796A-21942
22	7	1.8	77	2	US-09-248-796A-27087
23	7	1.8	87	2	US-09-248-796A-25044
24	7	1.8	90	2	US-09-613-486-24
25	7	1.8	91	2	US-09-902-540-12006
26	7	1.8	94	2	US-09-621-976-5665
27	7	1.8	112	2	US-09-543-681A-8319

28	7	1.8	120	2	US-09-311-021-20	Sequence 20, Appl
29	7	1.8	126	2	US-09-252-991A-28432	Sequence 28432, A
30	7	1.8	143	2	US-09-270-767-44294	Sequence 44294, A
31	7	1.8	152	2	US-09-248-796A-15608	Sequence 15608, A
32	7	1.8	160	2	US-09-543-681A-8310	Sequence 8310, Ap
33	7	1.8	178	2	US-09-489-039A-8236	Sequence 8236, Ap
34	7	1.8	182	2	US-09-489-039A-8301	Sequence 8301, Ap
35	7	1.8	212	2	US-09-270-767-41907	Sequence 41907, A
36	7	1.8	214	2	US-09-391-741A-2	Sequence 2, Appl
37	7	1.8	214	2	US-09-391-741A-16	Sequence 16, Appl
38	7	1.8	214	2	US-09-391-741A-26	Sequence 26, Appl
39	7	1.8	249	2	US-09-252-991A-30185	Sequence 30185, A
40	7	1.8	299	2	US-09-248-796A-15986	Sequence 15986, A
41	7	1.8	347	2	US-09-949-016-6928	Sequence 6928, Ap
42	7	1.8	352	2	US-09-073-009-14	Sequence 14, Appl
43	7	1.8	352	2	US-09-073-010-14	Sequence 14, Appl
44	7	1.8	365	2	US-09-270-767-46506	Sequence 46506, A
45	7	1.8	371	2	US-09-949-016-8425	Sequence 8425, Ap

ALIGNMENTS

RESULT 1
US-09-270-767-44680
; Sequence 44680, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 44680
; LENGTH: 401
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-44680

Query Match 2.3%; Score 9; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 244 SSQSSTQSK 252
Db 113 SSQSSTQSK 121

RESULT 2
US-09-513-999C-6360
; Sequence 6360, Application US/09513399C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.Pm
; SEQ ID NO 6360
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapiens

```
/ FEATURE:
/ NAME/KEY: UNSURE
/ LOCATION: 8
/ OTHER INFORMATION: Xaa=Leu or Pro
/ FEATURE:
/ NAME/KEY: UNSURE
/ LOCATION: 11
/ OTHER INFORMATION: Xaa=Glu or Lys
US-09-513-999C-6360

Query Match          2.0%; Score 8; DB 2; Length 55;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 251 SKEKKKK 258
Db 29 SKEKKKK 36

RESULT 3
US-09-149-476-615
/ Sequence 615, Application US/09149476
/ Patent No. 6420526
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: 186 Human Secreted proteins
/ FILE REFERENCE: PZ002P1
/ CURRENT APPLICATION NUMBER: US/09/149,476
/ CURRENT FILING DATE: 1998-09-08
/ EARLIER APPLICATION NUMBER: PCT/US98/04493
/ EARLIER FILING DATE: 1998-03-06
/ EARLIER APPLICATION NUMBER: 60/040,162
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,333
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/038,621
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,626
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,334
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,336
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/040,163
/ EARLIER FILING DATE: 1997-03-07
/ EARLIER APPLICATION NUMBER: 60/047,600
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,615
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,597
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,502
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,633
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,583
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,617
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,618
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,503
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,592
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,581
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,584
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,500
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,587
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,492
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,598
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,613
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,582
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,596
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,612
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,632
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/047,601
/ EARLIER FILING DATE: 1997-05-23
/ EARLIER APPLICATION NUMBER: 60/043,580
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,568
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,314
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,569
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,311
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,671
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,674
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,669
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,312
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,313
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,672
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/043,315
/ EARLIER FILING DATE: 1997-04-11
/ EARLIER APPLICATION NUMBER: 60/048,974
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/056,886
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,877
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,889
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,893
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,630
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,878
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,662
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,872
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,882
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,637
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,903
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,888
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,879
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,880
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,894
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,911
```

; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,636
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,874
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,910
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,864
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,631
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,845
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,892
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,761
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/047,595
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,599
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,588
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,585
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,586
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,590
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,594
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,589
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,593
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,614
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,578
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,576
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/047,501
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,670
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/056,632
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,664
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,876
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,881
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,909
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,875
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,862
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,887
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,908
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/057,650
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/056,884
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/057,669
; EARLIER FILING DATE: 1997-09-05
; EARLIER APPLICATION NUMBER: 60/049,610
; EARLIER FILING DATE: 1997-06-13

; EARLIER APPLICATION NUMBER: 60/061,060
; EARLIER FILING DATE: 1997-10-02

Query Match 2.0%; Score 8; DB 2; Length 61;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 244 SSSSTQS 251
Db 44 SSSSTQS 51
|||||||

RESULT 4
US-09-270-767-32008
; Sequence 32008, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of *Drosophila melanogaster*
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 32008
; LENGTH: 136
; TYPE: PRT
; ORGANISM: *Drosophila melanogaster*
US-09-270-767-32008

Query Match 2.0%; Score 8; DB 2; Length 136;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 331 HLETVIPA 338
Db 72 HLETVIPA 79
|||||||

RESULT 5
US-08-809-267-10
; Sequence 10, Application US/0809267
; Patent No. 5861296
; GENERAL INFORMATION:
; APPLICANT: LENNOX, Tricia L.
; APPLICANT: SLATKO, Lauren E.
; APPLICANT: SEARS, Lauren E.
; TITLE OF INVENTION: PURIFIED THERMOSTABLE INORGANIC
; TITLE OF INVENTION: PYROPHOSPHATASE OBTAINABLE FROM THERMOCOCCUS
; TITLE OF INVENTION: LITORALIS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEW ENGLAND BIOLABS, INC.
; STREET: 32 TOZER ROAD
; CITY: BEVERLY
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/809,267
; FILING DATE: 12-MAR-1997
; CLASSIFICATION: 433
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13662
; FILING DATE:
; APPLICATION NUMBER: US 08/329,721
; FILING DATE: 25-OCT-1994
; ATTORNEY/AGENT INFORMATION:

```
; NAME: WILLIAMS, GREGORY D.
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-105-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 927-5054
; TELEFAX: (508) 927-1705
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-809-267-10

Query Match      2.0%; Score 8; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      317 KMIDSGDK 324
Db      95 KMIDSGDK 102

RESULT 6
US-10-135-807-8
; Sequence 8, Application US/10135807
; Patent No. 6951744
; GENERAL INFORMATION:
; APPLICANT: The Secretary of State for Defence in Her Britannic Majesty's
; APPLICANT: Government of the United Kingdom of Great Britain and
; APPLICANT: No. 6951744thern Ireland
; APPLICANT: Clark, Duncan Roy
; APPLICANT: Vincent, Suzanne P
; TITLE OF INVENTION: Amplification process
; FILE REFERENCE: CG/P/133/WOD
; CURRENT APPLICATION NUMBER: US/10/135,807
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: GB 0110501.4
; PRIOR FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Thermococcus litoralis
US-10-135-807-8

Query Match      2.0%; Score 8; DB 2; Length 176;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      317 KMIDSGDK 324
Db      95 KMIDSGDK 102

RESULT 7
PCT-US95-13662A-10
; Sequence 10, Application PC/TUS9513662A
; GENERAL INFORMATION:
; APPLICANT: LENNOX, Tricia L.
; APPLICANT: SLATKO, Barton E.
; APPLICANT: SEARS, Lauren E.
; TITLE OF INVENTION: PURIFIED THERMOSTABLE INORGANIC
; TITLE OF INVENTION: PYROPHOSPHATASE OBTAINABLE FROM THERMOCOCCUS
; TITLE OF INVENTION: LITORALIS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEW ENGLAND BIOLABS, INC.
; STREET: 32 TOZER ROAD
; CITY: BEVERLY
; STATE: MASSACHUSETTS
; COUNTRY: USA

; NAME: WILLIAMS, GREGORY D.
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-105-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 927-5054
; TELEFAX: (508) 927-1705
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-13662A-10

Query Match      2.0%; Score 8; DB 4; Length 176;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      317 KMIDSGDK 324
Db      95 KMIDSGDK 102

RESULT 8
US-10-135-807-6
; Sequence 6, Application US/10135807
; Patent No. 6951744
; GENERAL INFORMATION:
; APPLICANT: The Secretary of State for Defence in Her Britannic Majesty's
; APPLICANT: Government of the United Kingdom of Great Britain and
; APPLICANT: No. 6951744thern Ireland
; APPLICANT: Clark, Duncan Roy
; APPLICANT: Vincent, Suzanne P
; TITLE OF INVENTION: Amplification process
; FILE REFERENCE: CG/P/133/WOD
; CURRENT APPLICATION NUMBER: US/10/135,807
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: GB 0110501.4
; PRIOR FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-135-807-6

Query Match      2.0%; Score 8; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      317 KMIDSGDK 324
Db      95 KMIDSGDK 102

RESULT 9
US-10-135-807-7
; Sequence 7, Application US/10135807
```

; Patent No. 6951744
; GENERAL INFORMATION:
; APPLICANT: The Secretary of State for Defence in Her Britannic Majesty's
; APPLICANT: Government of the United Kingdom of Great Britain and
; APPLICANT: No. 6951744thern Ireland
; APPLICANT: Clark, Duncan Roy
; APPLICANT: Vincent, Suzanne P
; TITLE OF INVENTION: Amplification process
; FILE REFERENCE: CG/P/133/WOD
; CURRENT APPLICATION NUMBER: US/10/135,807
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-10-135-807-7

Query Match 2.0%; Score 8; DB 2; Length 178;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 317 KMIDSGDK 324
Db 95 KMIDSGDK 102

RESULT 10
US-09-171-461-14
; Sequence 14, Application US/09171461
; Patent No. 6335016
; GENERAL INFORMATION:
; APPLICANT: Baker, Adam
; APPLICANT: Cotten, Matthew
; APPLICANT: Chiocca, Susanna
; APPLICANT: Kurzbaue, Robert
; APPLICANT: Schaffner, Gotthold
; TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus
; FILE REFERENCE: 0652.1800000
; CURRENT APPLICATION NUMBER: US/09/171,461
; CURRENT FILING DATE: 1999-01-12
; EARLIER APPLICATION NUMBER: PCT/EP97/01944
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 984
; TYPE: PRT
; ORGANISM: CELO Virus
; FEATURE:

; OTHER INFORMATION: Position: 23680..26634 /gene: L4 /product: L4 100K
US-09-171-461-14

Query Match 2.0%; Score 8; DB 2; Length 984;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 FLTPKAIA 13
Db 301 FLTPKAIA 308

RESULT 11
US-09-970-711-14
; Sequence 14, Application US/09970711
; Patent No. 6773709
; GENERAL INFORMATION:
; APPLICANT: Baker, Adam
; APPLICANT: Cotten, Matthew
; APPLICANT: Chiocca, Susanna

; APPLICANT: Kurzbaue, Robert
; APPLICANT: Schaffner, Gotthold
; TITLE OF INVENTION: Chicken Embryo Lethal Orphan (CELO) Virus
; FILE REFERENCE: 0652.1800001
; CURRENT APPLICATION NUMBER: US/09/970,711
; CURRENT FILING DATE: 2001-10-05
; PRIOR FILING DATE: 1999-01-12
; PRIOR FILING DATE: 1999-01-12
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 984
; TYPE: PRT
; ORGANISM: CELO Virus
; FEATURE:
; OTHER INFORMATION: Position: 23680..26634 /gene: L4 /product: L4 100K
US-09-970-711-14

Query Match 2.0%; Score 8; DB 2; Length 984;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 FLTPKAIA 13
Db 301 FLTPKAIA 308

RESULT 12
US-09-248-796A-22505
; Sequence 22505, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 22505
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-22505

Query Match 1.8%; Score 7; DB 2; Length 61;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 KEKKEKK 258
Db 4 KEKKEKK 10

RESULT 13
US-08-971-089-2
; Sequence 2, Application US/08971089
; Patent No. 6376174
; GENERAL INFORMATION:
; APPLICANT: Pulat, Stefan M.
; APPLICANT: Scoles, Daniel R.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING
; TITLE OF INVENTION: SCHWANNOMIN-BINDING-PROTEINS AND PRODUCTS RELATED THEREO
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: CAMPBELL & FLORES, LLP
; STREET: 4370 La Jolla Village Drive, Suite 700

```
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/971,089
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/030,987
; FILING DATE: 15-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ramos, Robert T.
; REGISTRATION NUMBER: 37,915
; REFERENCE/DOCKET NUMBER: P-CE 2862
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619)535-9001
; TELEFAX: (619)535-8949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-971-089-2

Query Match      1.8%; Score 7; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      252 KKKKKK 258
DB      14 KKKKKK 20

US-10-117-604A-2
; Sequence 2, Application US/10117604A
; Patent No. 6960650
; GENERAL INFORMATION:
; APPLICANT: Pulst, Stefan M.
; TITLE OF INVENTION: Nucleic Acid Encoding
; TITLE OF INVENTION: Schwannomin-Binding-Proteins and Products Related Thereto
; FILE REFERENCE: 66783-101
; CURRENT APPLICATION NUMBER: US/10/117,604A
; CURRENT FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/030,987
; PRIOR FILING DATE: 1996-11-15
; PRIOR APPLICATION NUMBER: US 08/971,089
; PRIOR FILING DATE: 1997-11-14
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-117-604A-2

Query Match      1.8%; Score 7; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      252 KKKKKK 258
DB      14 KKKKKK 20
```

Search completed: November 23, 2005, 16:50:49
Job time : 34.0842 secs

```
RESULT 14
US-09-513-999C-5320
; Sequence 5320 Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 5320
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-513-999C-5320
```

```
Query Match      1.8%; Score 7; DB 2; Length 63;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      252 KKKKKK 258
DB      46 KKKKKK 52
```

RESULT 15

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:44:35 ; Search time 111.283 Seconds
(without alignments)
1475.577 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 393
Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPLKGRVGEIQYEDISKLA 393

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1867569 seqs, 417829326 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA.Main.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pbp.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pbp.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pbp.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pbp.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pbp.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pbp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	393	100.0	393	4	US-10-408-765A-1465 Sequence 1465, Ap
2	73	18.6	84	4	US-10-106-698-5464 Sequence 5464, Ap
3	35	8.9	120	4	US-10-767-701-57151 Sequence 57151, A
4	35	8.9	138	4	US-10-437-963-191916 Sequence 191916, A
5	35	8.9	423	5	US-10-739-930-8499 Sequence 8499, Ap
6	35	8.9	424	4	US-10-425-115-341788 Sequence 341788, A
7	35	8.9	437	4	US-10-425-114-48183 Sequence 48183, A
8	35	8.9	437	4	US-10-425-114-58982 Sequence 58982, A
9	30	7.6	176	4	US-10-424-599-168326 Sequence 168326, A
10	15	3.8	390	6	US-10-097-143-41040 Sequence 41040, A
11	9	2.3	146	4	US-10-424-599-160516 Sequence 160516, A
12	9	2.3	374	4	US-10-424-599-264493 Sequence 264493, A
13	8	2.0	8	4	US-10-022-066-38 Sequence 38, Appl1
14	8	2.0	8	4	US-10-022-066-610 Sequence 610, App
15	8	2.0	39	4	US-10-425-115-357220 Sequence 357220, A
16	8	2.0	42	4	US-10-424-599-247175 Sequence 247175, A
17	8	2.0	61	3	US-09-809-391-615 Sequence 615, App
18	8	2.0	61	3	US-09-882-171-615 Sequence 615, App
19	8	2.0	63	4	US-10-164-861-615 Sequence 615, App
20	8	2.0	63	4	US-10-424-599-225902 Sequence 225902, A
21	8	2.0	64	4	US-10-425-115-264714 Sequence 264714, A
22	8	2.0	66	4	US-10-424-599-177043 Sequence 177043, A
23	8	2.0	72	4	US-10-424-599-173970 Sequence 173970, A
24	8	2.0	77	4	US-10-425-115-240939 Sequence 240939, A
25	8	2.0	82	4	US-10-425-115-230667 Sequence 230667, A
26	8	2.0	84	4	US-10-425-115-226277 Sequence 226277, A
27	8	2.0	89	4	US-10-425-115-281398 Sequence 281398, A

ALIGNMENTS

RESULT 1

US-10-408-765A-1465
; Sequence 1465, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary W.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1465
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1465

Query Match	100.0%;	Score 393;	DB 4;	Length 393;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 393;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGKSDFLTPKAIANRIKSGKGLKRLWYCMQCKQCRDENGFKCHCMSESHORQLLAS	60	Sequence 314787, A
Db	1	MGKSDFLTPKAIANRIKSGKGLKRLWYCMQCKQCRDENGFKCHCMSESHORQLLAS	60	Sequence 202154, A
Qy	61	PQPFMDYFSEFRNDFLELLRRRFGTKRVHNNIYVNEYISHREHNNATOWETLTDTFTK	120	Sequence 237652, A
Db	61	PQPFMDYFSEFRNDFLELLRRRFGTKRVHNNIYVNEYISHREHNNATOWETLTDTFTK	120	Sequence 190394, A
Qy	121	WLGREGUCKVDYTPKGYIYQIDRDPETIRRLQLEKKKQDLDDDEKTKAFIEQVRRG	180	Sequence 341154, A
Db	121	WLGREGUCKVDYTPKGYIYQIDRDPETIRRLQLEKKKQDLDDDEKTKAFIEQVRRG	180	Sequence 69213, A
Qy	181	LEGKEQVPTFTLSRENDEBEKVTFNLSKGACSSSGATSSKSTLGPSSALKTIGSSASVK	240	Sequence 45547, A
Db	181	LEGKEQVPTFTLSRENDEBEKVTFNLSKGACSSSGATSSKSTLGPSSALKTIGSSASVK	240	Sequence 287597, A
Qy	241	RKSSQSSTOSKEKKKKKSSALDEIMEIEBEKRTARTDYWLOPEIIIVKIIITKLGKGYHK	300	Sequence 6, Appl1
Db	241	RKSSQSSTOSKEKKKKKSSALDEIMEIEBEKRTARTDYWLOPEIIIVKIIITKLGKGYHK	300	Sequence 7, Appl1
Qy	301	KKAIVKEVIDKYTAVVQWMDSGDKLKDQTHLETVPAPGKRILVLNGVGRNGEGTLES	360	Sequence 340712, A

Db 301 KKAIVKEVDKTVAVVKMIDSGDKLQDTHLETVPAPGKRILVNLGGYRNGEFTLESI 360
QY 361 NEKTSATIVETGPKLGRRVEGIOYEDISKLA 393
Db 361 NEKTSATIVETGPKLGRRVEGIOYEDISKLA 393
RESULT 2
US-10-106-698-5464
; Sequence 5464, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5464
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5464
Query Match 18.6%; Score 73; DB 4; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.3e-59;
Matches 73; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 321 SGDKLQDTHLETVPAPGKRILVNLGGYRNGEFTLESI NEKTSATIVETGPKLGR 380
Db 12 SGDKLQDTHLETVPAPGKRILVNLGGYRNGEFTLESI NEKTSATIVETGPKLGR 71
QY 381 VEGIOYEDISKLA 393
Db 72 VEGIOYEDISKLA 84
RESULT 3
US-10-767-701-57151
; Sequence 57151, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 57151
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30963818.p
US-10-767-701-57151
Query Match 8.9%; Score 35; DB 4; Length 120;
Best Local Similarity 100.0%; Pred. No. 4.2e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHORQ 53

Db 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHORQ 53
RESULT 4
US-10-437-963-191916
; Sequence 191916, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 191916
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_88190C.1.pap
US-10-437-963-191916
Query Match 8.9%; Score 35; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 4.7e-24;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHORQ 53
Db 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHORQ 53
RESULT 5
US-10-739-930-8499
; Sequence 8499, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 8499
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEAMA-23APR03-C7583_2.p
US-10-739-930-8499
Query Match 8.9%; Score 35; DB 5; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.2e-23;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHORQ 53
Db 19 KGLQKLRYWCQKQCRDENGFKCHCMSESHORQ 53
RESULT 6
US-10-425-115-341788
; Sequence 341788, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:


```
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 341788
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_74879C.1.pep
US-10-425-115-341788

Query Match      8.9%; Score 35; DB 4; Length 424;
Best Local Similarity 100.0%; Pred. No. 1.2e-23;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYCQMCQKQCRDENGFKCHCMSESHQRQ 53
    |||||
DB 19 KGLQKLRYCQMCQKQCRDENGFKCHCMSESHQRQ 53
    |||||

RESULT 7
US-10-425-114-48183
; Sequence 48183, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48183
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-011-E3_FLI.pep
US-10-425-114-48183

Query Match      8.9%; Score 35; DB 4; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.3e-23;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYCQMCQKQCRDENGFKCHCMSESHQRQ 53
    |||||
DB 32 KGLQKLRYCQMCQKQCRDENGFKCHCMSESHQRQ 66
    |||||

RESULT 8
US-10-425-114-58982
; Sequence 58982, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
```

```
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 58982
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700238628_FLI.pep
US-10-425-114-58982

Query Match      8.9%; Score 35; DB 4; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.3e-23;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYCQMCQKQCRDENGFKCHCMSESHQRQ 53
    |||||
DB 32 KGLQKLRYCQMCQKQCRDENGFKCHCMSESHQRQ 66
    |||||

RESULT 9
US-10-424-599-168326
; Sequence 168326, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 168326
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_123013C.1.pep
US-10-424-599-168326

Query Match      7.6%; Score 30; DB 4; Length 176;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGLQKLRYCQMCQKQCRDENGFKCHCMSE 48
    |||||
DB 19 KGLQKLRYCQMCQKQCRDENGFKCHCMSE 48
    |||||

RESULT 10
US-11-097-143-41040
; Sequence 41040, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
```

```
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41040
; LENGTH: 390
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-41040

Query Match          3.8%; Score 15; DB 6; Length 390;
Best Local Similarity 100.0%; Pred. No. 4.8e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 KSKGLQKLRWYCMC 31
Db 17 KSKGLQKLRWYCMC 31

RESULT 11
US-10-424-599-160516
; Sequence 160516, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 160516
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_115964C.1.pep
US-10-424-599-160516

Query Match          2.3%; Score 9; DB 4; Length 146;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 SKEKKKKKS 259
Db 56 SKEKKKKKS 64

RESULT 12
US-10-424-599-264493
; Sequence 264493, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
```

```
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 264493
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(374)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_80858C.1.pep
US-10-424-599-264493

Query Match          2.3%; Score 9; DB 4; Length 374;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 251 SKEKKKKKS 259
Db 124 SKEKKKKKS 132

RESULT 13
US-10-022-066-38
; Sequence 38, Application US/10022066
; Publication No. US20030166057A1
; GENERAL INFORMATION:
; APPLICANT: HILDEBRAND, WILLIAM H.
; APPLICANT: PRILLMAN, KILEY RAE
; TITLE OF INVENTION: METHOD AND APPARATUS FOR THE PRODUCTION OF ANTIGENS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 6680.034
; CURRENT APPLICATION NUMBER: US/10/022,066
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/256,410
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/256,409
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 09/465,321
; PRIOR FILING DATE: 1999-12-17
; PRIOR APPLICATION NUMBER: 09/974,366
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 38
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-10-022-066-38

Query Match          2.0%; Score 8; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 SHORQLLL 56
Db 1 SHORQLLL 8

RESULT 14
US-10-022-066-610
; Sequence 610, Application US/10022066
; Publication No. US20030166057A1
; GENERAL INFORMATION:
; APPLICANT: HILDEBRAND, WILLIAM H.
; APPLICANT: PRILLMAN, KILEY RAE
; TITLE OF INVENTION: METHOD AND APPARATUS FOR THE PRODUCTION OF ANTIGENS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: 6680.034
; CURRENT APPLICATION NUMBER: US/10/022,066
```

Wed Nov 30 09:40:51 2005

```
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/256,410
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 60/256,409
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: 09/465,321
; PRIOR FILING DATE: 1999-12-17
; PRIOR APPLICATION NUMBER: 09/974,366
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 638
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 610
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide
US-10-022-066-610

Query Match          2.0%; Score 8; DB 4; Length 8;
Best Local Similarity 100.0%; Pred. No. 1.7e+06;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 SHORQLLL 56
Db 1 SHORQLLL 8

RESULT 15
US-10-425-115-357220
; Sequence 357220, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 357220
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_88955C.1.pep
US-10-425-115-357220

Query Match          2.0%; Score 8; DB 4; Length 39;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 KERKKKKKS 259
Db 31 KERKKKKKS 38

Search completed: November 23, 2005, 16:55:49
Job time : 112.283 secs
```

This Page Blank (usp1c)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:40:00 ; Search time 5.01276 Seconds
(without alignments)
237.628 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 393
Sequence: 1 MGKSDITPKAIANRIKSG.....GPLKGRVGEIQYEDISKLA 393

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 17545 seqs, 3030971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 17545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA New:
1: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.0	120	1	US-10-821-234-1011
2	6	1.5	27	1	US-10-952-535A-31
3	6	1.5	57	1	US-10-723-207-88
4	6	1.5	112	1	US-10-793-626-2114
5	6	1.5	133	1	US-10-667-295-49
6	6	1.5	149	1	US-10-986-501-231
7	6	1.5	175	1	US-10-667-295-48
8	6	1.5	176	1	US-10-507-275-10
9	6	1.5	204	1	US-10-793-626-2394
10	6	1.5	226	1	US-10-485-517-186
11	6	1.5	227	1	US-10-986-501-136
12	6	1.5	227	1	US-10-793-626-2642
13	6	1.5	229	7	US-11-014-629-1
14	6	1.5	258	7	US-11-038-284-23
15	6	1.5	286	7	US-10-667-295-60
16	6	1.5	286	7	US-11-137-465-48
17	6	1.5	327	1	US-10-667-295-59
18	6	1.5	339	1	US-10-821-234-1507
19	6	1.5	389	7	US-11-012-762-72
20	6	1.5	398	7	US-11-012-762-74
21	6	1.5	441	1	US-10-510-386-162
22	6	1.5	458	7	US-11-077-550-114
23	6	1.5	467	1	US-10-510-386-56
24	6	1.5	471	1	US-10-467-962B-10
25	6	1.5	548	7	US-11-077-550-24

26	6	1.5	572	1	US-10-821-234-1290	Sequence 1290, Ap
27	6	1.5	600	1	US-10-131-826A-462	Sequence 462, Ap
28	6	1.5	674	1	US-10-507-275-9	Sequence 9, Appl
29	6	1.5	724	1	US-10-793-626-968	Sequence 968, App
30	6	1.5	749	1	US-10-793-626-138	Sequence 138, App
31	6	1.5	853	1	US-10-821-234-1110	Sequence 1110, Ap
32	6	1.5	858	7	US-11-077-550-22	Sequence 22, Appl
33	6	1.5	860	7	US-11-077-550-175	Sequence 175, App
34	6	1.5	862	7	US-11-077-550-94	Sequence 94, Appl
35	6	1.5	862	7	US-11-077-550-171	Sequence 171, App
36	6	1.5	862	7	US-11-077-550-173	Sequence 173, App
37	6	1.5	864	7	US-11-077-550-102	Sequence 102, App
38	6	1.5	865	7	US-11-077-550-100	Sequence 100, App
39	6	1.5	866	7	US-11-077-550-88	Sequence 88, Appl
40	6	1.5	866	7	US-11-077-550-104	Sequence 104, App
41	6	1.5	867	7	US-11-077-550-80	Sequence 80, Appl
42	6	1.5	867	7	US-11-077-550-96	Sequence 96, Appl
43	6	1.5	867	7	US-11-077-550-98	Sequence 98, Appl
44	6	1.5	870	7	US-11-077-550-92	Sequence 92, Appl
45	6	1.5	871	7	US-11-077-550-84	Sequence 84, Appl

ALIGNMENTS

RESULT 1

US-10-821-234-1011
; Sequence 1011, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1011
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1011

Query Match 2.0%; Score 8; DB 1; Length 120;
Best Local Similarity 100.0%; Pred. No. 0.34;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 252 KEKKEKKS 259
Db 15 KEKKEKKS 22

RESULT 2

US-10-952-535A-31
; Sequence 31, Application US/10952535A
; Publication No. US20050255113A1
; GENERAL INFORMATION:
; APPLICANT: Huston, James S.
; APPLICANT: Messer, Anne
; APPLICANT: Lecerif, Jean-Michel
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING POLYPEPTIDE-
; FILE REFERENCE: INR-004CP
; TITLE OF INVENTION: ACCUMULATION ASSOCIATED WITH NEUROLOGICAL DISORDERS
; CURRENT APPLICATION NUMBER: US/10/952,535A
; CURRENT FILING DATE: 2004-09-27
; PRIOR APPLICATION NUMBER: 60/146,047
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 45

```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: construct
US-10-952-535A-31

Query Match          1.5%; Score 6; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 KKKKK 259
Db 15 KKKKK 20

RESULT 3
US-10-723-207-88
; Sequence 88, Application US/10723207
; Publication No. US20050250934A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Chang Yi
; TITLE OF INVENTION: PEPTIDE COMPOSITION AS IMMUNOGEN FOR THE TREATMENT OF
; TITLE OF INVENTION: ALLERGY
; FILE REFERENCE: 1151-4153US2
; CURRENT APPLICATION NUMBER: US/10/723,207
; CURRENT FILING DATE: 2003-11-24
; PRIOR APPLICATION NUMBER: 09/701,623
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US99/13959
; PRIOR FILING DATE: 1999-06-21
; PRIOR APPLICATION NUMBER: 09/100,287
; PRIOR FILING DATE: 1998-06-20
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 88
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Peptide
; OTHER INFORMATION: synthesized from amino acids with no genetic
; OTHER INFORMATION: material as source
US-10-723-207-88

Query Match          1.5%; Score 6; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 KKKKK 257
Db 11 KKKKK 16

RESULT 4
US-10-793-626-2114
; Sequence 2114, Application US/10793626
; Publication No. US2005025478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2114
```

```
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (112)
; OTHER INFORMATION: variable amino acid
US-10-793-626-2114

Query Match          1.5%; Score 6; DB 1; Length 112;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 QRQLLL 56
Db 35 QRQLLL 40

RESULT 5
US-10-667-295-49
; Sequence 49, Application US/10667295
; Publication No. US20050257293A1
; GENERAL INFORMATION:
; APPLICANT: Mascia, Peter
; TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
; FILE REFERENCE: 11696-047001
; CURRENT APPLICATION NUMBER: US/10/667,295
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: US 60/411,823
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 263
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Brassica napus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(133)
; OTHER INFORMATION: Ceres Seq. ID no. 6416839
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(133)
; OTHER INFORMATION: Xaa = any amino acid
US-10-667-295-49

Query Match          1.5%; Score 6; DB 1; Length 133;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 SSASVK 240
Db 18 SSASVK 23

RESULT 6
US-10-986-501-231
; Sequence 231, Application US/10986501
; Publication No. US20050244845A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P2C1
; CURRENT APPLICATION NUMBER: US/10/986,501
; CURRENT FILING DATE: 2004-11-12
; PRIOR APPLICATION NUMBER: US/10/621,363
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: 09/969,730
; PRIOR FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: 09/774,639
```

```
; PRIOR FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; PRIOR FILING DATE: 1997-08-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 231
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-986-501-231

Query Match      1.5%; Score 6; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 LEKKKK 160
DB 141 LEKKKK 146
|||||

RESULT 7
US-10-667-295-48
; Sequence 48, Application US/10667295
; Publication No. US20050257293A1
; GENERAL INFORMATION:
; APPLICANT: Masclia, Peter
; TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
; FILE REFERENCE: 11696-047001
; CURRENT APPLICATION NUMBER: US/10/667,295
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: US 60/411,823
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 263
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Brassica napus
; NAME/KEY: VARIANT
; LOCATION: (1)...(175)
; FEATURE:
; OTHER INFORMATION: Ceres Seq. ID no. 6416838
; NAME/KEY: VARIANT
; LOCATION: (1)...(175)
; OTHER INFORMATION: Xaa = any amino acid
US-10-667-295-48

Query Match      1.5%; Score 6; DB 1; Length 175;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 SSASVK 240
DB 60 SSASVK 65
|||||

RESULT 8
US-10-507-275-10
; Sequence 10, Application US/10507275
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; Publication No. US20050250166A1
; GENERAL INFORMATION:
; APPLICANT: Masai, Hisao
; APPLICANT: Tamai, Katsuyuki
; APPLICANT: Medical and Biological Laboratories Co., Ltd.
; APPLICANT: Japan Science and Technology Agency
; APPLICANT: Ginkgo Biomedical Research Institute Co., Ltd.
; TITLE OF INVENTION: Cdc7-ASK Kinase Complex, Substrates of the Kinase Complex,
; TITLE OF INVENTION: Specific Antibodies to the Substrates, and Screening Methods
; TITLE OF INVENTION: Using the Same to Screen for Compounds Comprising Cdc7-ASK
; TITLE OF INVENTION: Kinase Inhibitory Ability
; FILE REFERENCE: 082368-001100US
; CURRENT APPLICATION NUMBER: US/10/507,275
; CURRENT FILING DATE: 2004-09-09
; PRIOR APPLICATION NUMBER: JP 2002-067702
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: WO PCT/JP03/02918
; PRIOR FILING DATE: 2003-03-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-507-275-10

Query Match      1.5%; Score 6; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 KEKKKK 257
DB 115 KEKKKK 120
|||||

RESULT 9
US-10-793-626-2394
; Sequence 2394, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2394
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2394

Query Match      1.5%; Score 6; DB 1; Length 204;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 QRQLLL 56
DB 35 QRQLLL 40
|||||

RESULT 10
US-10-485-517-186
; Sequence 186, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
```

```
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118925.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 186
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-186

Query Match          1.5%; Score 6; DB 1; Length 226;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 TIGSSA 237
DB 124 TIGSSA 129

RESULT 11
US-10-986-501-136
; Sequence 136, Application US/10986501
; Publication No. US20050244845A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P2C1
; CURRENT APPLICATION NUMBER: US/10/986,501
; CURRENT FILING DATE: 2004-11-12
; PRIOR APPLICATION NUMBER: US/10/621,363
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: 09/969,730
; PRIOR FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: 09/774,639
; PRIOR FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; PRIOR FILING DATE: 1997-08-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 136
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-986-501-136

Query Match          1.5%; Score 6; DB 1; Length 227;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 EKKKKK 258
DB 124 EKKKKK 129

; APPLICANT: Kimmerly, William John
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUS3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2642
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2642

Query Match          1.5%; Score 6; DB 1; Length 227;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 355 GTLESI 360
DB 57 GTLESI 62

RESULT 13
US-11-014-629-1
; Sequence 1, Application US/11014629
; Publication No. US20050244376A1
; GENERAL INFORMATION:
; APPLICANT: Hall, Frederick L.
; APPLICANT: Gordon, Erlinda M.
; TITLE OF INVENTION: TARGETED GENE DELIVERY TO EXTRACELLULAR MATRIX
; FILE REFERENCE: 30863-704.302
; CURRENT APPLICATION NUMBER: US/11/014,629
; CURRENT FILING DATE: 2004-12-15
; PRIOR APPLICATION NUMBER: US 08/837,223
; PRIOR FILING DATE: 1997-04-10
; PRIOR APPLICATION NUMBER: US 09/904,923
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 229
; TYPE: PRT
; ORGANISM: moloney murine leukemia virus
US-11-014-629-1

Query Match          1.5%; Score 6; DB 7; Length 229;
Best Local Similarity 100.0%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 LKLDQT 330
DB 103 LKLDQT 108

RESULT 14
US-11-038-284-23
; Sequence 23, Application US/11038284
; Publication No. US20050246793A1
; GENERAL INFORMATION:
; APPLICANT: COOKE, DAVID
```



```

; APPLICANT: DEBET, MARTINE
; APPLICANT: GIDLEY, MICHAEL, JOHN
; APPLICANT: JOBLING, STEPHEN, ALAN
; APPLICANT: SAFFORD, RICHARD
; APPLICANT: SIDEBOTTOM, CHRISTOPHER, MICHAEL
; APPLICANT: WESTCOTT, ROGER, JOHN
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO PLANT STARCH COMPOSITION
; FILE REFERENCE: 054163-5003-US
; CURRENT APPLICATION NUMBER: US/11/038,284
; CURRENT FILING DATE: 2005-01-21
; PRIOR APPLICATION NUMBER: US/10/056,454
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: PCT/GB96/01075
; PRIOR FILING DATE: 1996-05-03
; PRIOR APPLICATION NUMBER: GB 9607409.1
; PRIOR FILING DATE: 1996-04-10
; PRIOR APPLICATION NUMBER: GB 9509229.2
; PRIOR FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Zea mays
US-11-038-284-23

```

```

Query Match          1.5%; Score 6; DB 7; Length 258;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 198 NDEEKV 203
Db 101 NDEEKV 106

```

```

RESULT 15
US-10-667-295-60
; Sequence 60, Application US/10667295
; Publication No. US20050257293A1
; GENERAL INFORMATION:
; APPLICANT: Mascia, Peter
; TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
; FILE REFERENCE: 11696.047001
; CURRENT APPLICATION NUMBER: US/10/667,295
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: US 60/411,823
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 263
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 286
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(286)
; OTHER INFORMATION: Ceres Seq. ID no. 13491862
US-10-667-295-60

```

```

Query Match          1.5%; Score 6; DB 1; Length 286;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 235 SSASVK 240
Db 21 SSASVK 26

```

```

Search completed: November 23, 2005, 16:52:01
Job time : 5.01276 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 16:39:45 ; Search time 25.0638 Seconds
(without alignments)
1508.679 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 393
Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPLKGRVVEIGQYEDISKLA 393

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 80.*

- 1: Pir1.*
- 2: Pir2.*
- 3: Pir3.*
- 4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163	41.5	391	2	S18666
2	35	8.9	411	2	H96596
3	10	2.5	404	2	T27106
4	9	2.3	304	2	T40316
5	8	2.0	178	2	G75018
6	8	2.0	178	2	A1205
7	8	2.0	303	1	S75983
8	8	2.0	521	2	E64181
9	8	2.0	646	2	F71620
10	8	2.0	1799	2	A1895
11	7	1.8	57	2	A12525
12	7	1.8	68	2	E31176
13	7	1.8	70	2	E81029
14	7	1.8	89	2	AF2709
15	7	1.8	89	2	G37491
16	7	1.8	94	2	F86022
17	7	1.8	100	2	D82245
18	7	1.8	106	2	S59536
19	7	1.8	114	2	S64961
20	7	1.8	115	2	T33703
21	7	1.8	121	2	D75089
22	7	1.8	132	2	T49635
23	7	1.8	133	2	T22262
24	7	1.8	136	2	AB2542
25	7	1.8	137	2	S37848
26	7	1.8	138	2	B84793
27	7	1.8	138	2	T45905
28	7	1.8	144	2	G65255
29	7	1.8	150	2	H82341

30	7	1.8	153	2	S59587
31	7	1.8	153	2	S59125
32	7	1.8	153	2	S59591
33	7	1.8	175	2	T05669
34	7	1.8	177	2	PS0374
35	7	1.8	206	2	T25384
36	7	1.8	214	2	JC7297
37	7	1.8	215	2	AB3358
38	7	1.8	217	2	JC7997
39	7	1.8	217	2	T27198
40	7	1.8	239	2	C84994
41	7	1.8	239	2	AF0780
42	7	1.8	250	2	T44896
43	7	1.8	250	2	D70787
44	7	1.8	253	2	A05283
45	7	1.8	266	2	A86288

ALIGNMENTS

RESULT 1

S18666
KIN17 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S18666; A43753; S14622
R:Angulo, J.F.; Rouer, E.; Marin, A.; Mattei, M.G.; Tissier, A.; Horellou, P.; Benarous, R.
Nucleic Acids Res. 19, 5117-5123, 1991
A:Title: Identification and expression of the cDNA of KIN17, a zinc-finger gene located .
A:Reference number: S18666; MUID:92020193; PMID:1923796
A:Accession: S18666
A:Molecule type: mRNA
A:Residues: 1-391 <ANG>
A:Cross-references: UNIPROT:Q8K339; UNIPARC:UPI0000017993F; EMBL:X58472
R:Angulo, J.; Rouer, E.; Benarous, R.; Devoret, R.
Biochimie 73, 251-256, 1991
A:Title: Identification of a mouse cDNA fragment whose expressed polypeptide reacts with
A:Reference number: A43753; MUID:91355299; PMID:1715759
A:Accession: A43753
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 74-273 <ANW>
A:Cross-references: UNIPARC:UPI00000179940
C:Superfamily: KIN17 protein

Query Match		41.5%;	Score 163;	DB 2;	Length 391;
Best Local Similarity		100.0%;	Pred. No. 1.5e-153;		
Matches	163;	Conservative	0;	Mismatches	0;
		Indels		0;	Gaps 0;
Qy	25	RWYCOMCOKCRDENGFKCHMSHQRQLLASENQQPMFYSEFRNDFLLELRFF	84		
Db	25	RWYCOMCOKCRDENGFKCHMSHQRQLLASENQQPMFYSEFRNDFLLELRFF	84		
Qy	85	GTKVHNNIVYVNIYSHREIHNNATQWETLDTFKWLGREGLCCKVDTPKGVYQYIDR	144		
Db	85	GTKVHNNIVYVNIYSHREIHNNATQWETLDTFKWLGREGLCCKVDTPKGVYQYIDR	144		
Qy	145	DPETIRQLELEKKKKQDLDEKTKAFIEQVRRGLEKGEQ	187		
Db	145	DPETIRQLELEKKKKQDLDEKTKAFIEQVRRGLEKGEQ	187		

RESULT 2

H96596
hypothetical protein T5A14.13 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H96596
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H96596
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-411 <STO>
A;Cross-references: UNIPROT:Q9ZVU5; UNIPARC:UPI00000A3CC8; GB:AE005173; NID:g4204268; P
C;Genetics:
A;Gene: TSA14.13
A;Map position: 1
C;Superfamily: KIN17 protein

Query Match 8.9%; Score 35; DB 2; Length 411;
Best Local Similarity 100.0%; Pred. No. 2.2e-26;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 KGIQKLRLWYQCMQKQCRDENGFKCHCMSESHQRQ 53
|||||
Db 19 KGIQKLRLWYQCMQKQCRDENGFKCHCMSESHQRQ 53

RESULT 3
T27106
hypothetical protein Y52B11A.9 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T27106
R;Lennard, N.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z20310
A;Accession: T27106
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-404 <WIL>
A;Cross-references: UNIPROT:Q9XWF2; UNIPARC:UPI0000078B04; EMBL:AL032654; PIDN:CAA21720.
A;Experimental source: clone Y52B11A
C;Genetics:
A;Gene: CESP:Y52B11A.9
A;Map position: 1
A;Intons: 26/2; 102/3; 301/3; 367/3
C;Superfamily: KIN17 protein

Query Match 2.5%; Score 10; DB 2; Length 404;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 CQMCQKQCRD 37
|||||
Db 28 CQMCQKQCRD 37

RESULT 4
T40316
conserved kin7-like hypothetical protein SPBC365.09c - fission yeast (*Schizosaccharomyce*
C;Species: *Schizosaccharomyces pombe*
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T40316
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Lelaure, V.; Galibert, F.
submitted to the EMBL Data Library, June 1999
A;Reference number: Z21920
A;Accession: T40316
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-304 <WOO>
A;Cross-references: UNIPROT:Q9V7X9; UNIPARC:UPI000006B457; EMBL:AL078627; PIDN:CAB44761.
A;Experimental source: strain 972h-; cosmid c365
C;Genetics:
A;Gene: SPDB:SPBC365.09c

A;Map position: 2

Query Match 2.3%; Score 9; DB 2; Length 304;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 RDENGFKCH 44
|||||Db 36 RDENGFKCH 44
|||||

RESULT 5

F75018

inorganic diphosphatase (EC 3.6.1.1) PAB1104 [similarity] - *Pyrococcus abyssi* (strain Orf
C;Species: *Pyrococcus abyssi*

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004

C;Accession: F75018

R;anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A;Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome stru

A;Reference number: A75001

A;Accession: F75018

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-178 <KAW>

A;Cross-references: UNIPROT:Q9UY24; UNIPARC:UPI000003456C; GB:AJ248288; GB:AL096836; NID:

A;Experimental source: strain Orsay

C;Genetics:

A;Gene: ppa; PAB1104

A;Superfamily: inorganic pyrophosphatase

C;Keywords: hydrolase

Query Match 2.0%; Score 8; DB 2; Length 178;

Best Local Similarity 100.0%; Pred. No. 7.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 KNIDSGDK 324
|||||

Db 95 KNIDSGDK 102

RESULT 6

A71205

inorganic diphosphatase (EC 3.6.1.1) PH1907 [similarity] - *Pyrococcus horikoshii*C;Species: *Pyrococcus horikoshii*

C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Oct-2004

C;Accession: A71205

R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekiri

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,

DNA Res. 5, 55-76, 1998

A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a

A;Reference number: A71000; MUID:98344137; PMID:9679194

A;Accession: A71205

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-178 <KAW>

A;Cross-references: UNIPROT:O59570; UNIPARC:UPI0000052CB5; GB:AP000007; NID:g3236134; Pfi

A;Experimental source: strain Ors

A;Note: this accession replaces an interim accession for a sequence replaced by GenBank

C;Genetics:

A;Gene: PH1907

C;Superfamily: inorganic pyrophosphatase

C;Keywords: hydrolase

Query Match 2.0%; Score 8; DB 2; Length 178;

Best Local Similarity 100.0%; Pred. No. 7.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 317 KNIDSGDK 324
|||||

Db 95 KNIDSGDK 102

```

RESULT 7
S75983
hypothetical protein - Synchocystis sp. (strain PCC 6803)
C:Species: Synchocystis sp.
A:Variety: PCC 6803
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S75983
R;Kaneko, T.; Sato, S.; Kōtani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S75983
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <KAN>
A:Cross-references: UNIPROT:Q55483; UNIPARC:UPI000000COEFA; EMBL:D64006; GB:AB001339; NID
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: hypothetical protein ybbp

Query Match 2.0%; Score 8; DB 1; Length 303;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 76 FLELLRRR 83
|||||
Db 100 FLELLRRR 107

RESULT 8
E64181
probable cytochrome oxidase (EC 1.10.3.-) chain I - Haemophilus influenzae (strain Rd KW
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004
C:Accession: E64181
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: E64181
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-521 <TIGR>
A:Cross-references: UNIPROT:P45021; UNIPARC:UPI0000128C53; GB:U32787; GB:L42023; NID:G15
C:Superfamily: cytochrome d complex terminal oxidase chain I
C:Keywords: electron transfer; heme; oxidoreductase; transmembrane protein
F;20-36/Domain: transmembrane #status predicted <TM1>
F;100-116/Domain: transmembrane #status predicted <TM2>
F;192-208/Domain: transmembrane #status predicted <TM3>
F;222-238/Domain: transmembrane #status predicted <TM4>
F;398-414/Domain: transmembrane #status predicted <TM5>
F;426-442/Domain: transmembrane #status predicted <TM6>
F;477-493/Domain: transmembrane #status predicted <TM7>

Query Match 2.0%; Score 8; DB 2; Length 521;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 225 LGPSALKT 232
|||||
Db 504 LGPSALKT 511

RESULT 9
F71620
hypothetical protein PFB0230c - malaria parasite (Plasmodium falciparum)
C:Species: plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C:Accession: F71620

```

```

R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.)
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: F71620
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-646 <GAR>
A:Cross-references: UNIPROT:O96147; UNIPARC:UPI000017B603; GB:AE001381; GB:AE001362; NID:
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0230c

Query Match 2.0%; Score 8; DB 2; Length 646;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 251 SKEKKKKK 258
|||||
Db 575 SKEKKKKK 582

RESULT 10
AD1895
serine/threonine kinase with two-component sensor domain alr0710 [imported] - Nostoc sp;
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AD1895
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1895
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1799 <KUR>
A:Cross-references: UNIPROT:O8YVY2; UNIPARC:UPI000000CDB39; GB:BA000019; PIDN:BA073667.1)
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0710

Query Match 2.0%; Score 8; DB 2; Length 1799;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 213 SSSGATSS 220
|||||
Db 1325 SSSGATSS 1332

RESULT 11
AI2525
transposase asr7385 [imported] - Nostoc sp. (strain PCC 7120) plasmid pcc7120alpha
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AI2525
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AI2525
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-57 <KUR>
A:Cross-references: UNIPROT:O9YKB6; UNIPARC:UPI000000CEFC; GB:BA000020; PIDN:BA077143.1)
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: asr7385

```

```

A:Genome: plasmid

Query Match      1.8%; Score 7; DB 2; Length 57;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 LLRRRFG 85
    |||||
Db 40 LLRRRFG 46

RESULT 12
E91176
hypothetical protein ECs4381 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E91176
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E91176
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-68 <HAY>
A:Cross-references: UNIPROT:Q8X2W1; UNIPARC:UPI00000D2A54; GB:BA0000007; PIDN:BA037804.1;
A:Experimental source: strain O157:H7, substrain R1MD 050952
C:Genetics:
A:Gene: ECs4381

Query Match      1.8%; Score 7; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 OSKEKKK 256
    |||||
Db 2 OSKEKKK 8

RESULT 13
E81029
hypothetical protein NMB1892 [imported] - Neisseria meningitidis (strain MC58 serogroup
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C:Accession: E81029
R:Tetzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: E81029
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-70 <TET>
A:Cross-references: UNIPROT:Q9JXT4; UNIPARC:UPI00000C4805; GB:AE002539; GB:AE002098; NID
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1892

Query Match      1.8%; Score 7; DB 2; Length 70;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 RRGLECK 184
    |||||
Db 53 RRGLECK 59

RESULT 14
AF2709
hypothetical protein Atu1079 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AF2709
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AF2709
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-89 <KUR>
A:Cross-references: UNIPROT:Q8UGF8; UNIPARC:UPI00000D1A2B; GB:AE008688; PIDN:AAL42092.1;
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu1079
A:Map position: circular chromosome

Query Match      1.8%; Score 7; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 KFIERQV 177
    |||||
Db 12 KFIERQV 18

RESULT 15
G97491
hypothetical protein AGR_C_1994 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: G97491
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: G97491
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-89 <KUR>
A:Cross-references: UNIPROT:Q8UGF8; UNIPARC:UPI00000D1A2B; GB:AE007869; PIDN:AAK86888.1;
A:Gene: AGR_C_1994
A:Map position: circular chromosome

Query Match      1.8%; Score 7; DB 2; Length 89;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 KFIERQV 177
    |||||
Db 12 KFIERQV 18

Search completed: November 23, 2005, 16:51:47
Job time : 26.0638 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: November 23, 2005, 16:19:54 ; Search time 25.0638 Seconds
(without alignments)
1508.679 Million cell updates/sec

Title: US-09-555-529-26
Perfect score: 2037
Sequence: 1 MGKSDFLTPKAIANRIKSG.....GPLKGRVGEIQYEDISKLA 393

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80.*
1: Pirl.*
2: Pirl2.*
3: Pirl3.*
4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1887	92.6	391	2	S18666
2	978	48.0	411	2	H96596
3	952.5	46.8	404	2	T27106
4	586.5	28.8	304	2	T40316
5	300.5	14.8	232	2	S50239
6	165	8.1	462	2	A86459
7	143	7.0	482	2	T24518
8	141.5	6.9	517	1	A39038
9	140	6.9	375	2	T05523
10	139.5	6.8	764	2	A84671
11	138	6.8	805	2	E70474
12	133.5	6.6	630	2	H96770
13	133	6.5	1359	2	T34036
14	131.5	6.5	1877	2	T21861
15	129.5	6.4	2401	2	T28676
16	128	6.3	741	2	S12921
17	127.5	6.3	2541	2	T29340
18	127	6.2	2094	2	S33124
19	127	6.2	3488	2	T34418
20	126.5	6.2	485	2	S44819
21	126.5	6.2	1827	2	T16270
22	126	6.2	671	2	H64502
23	126	6.2	888	2	E71608
24	125.5	6.2	978	2	A70387
25	125	6.1	1188	2	E89896
26	123.5	6.1	1359	2	T52091
27	123.5	6.1	1388	2	T38720
28	123	6.0	377	2	S52537
29	123	6.0	1164	2	T24806

ALIGNMENTS

RESULT 1

S18666

KIN17 protein - mouse

C;Species: Mus musculus (house mouse)

C;Date: 19-Feb-1994 #sequence revision 10-Nov-1995 #text_change 09-Jul-2004

C;Accession: S18666; A43753; S14622

R;Angulo, J.F.; Rouer, E.; Mazin, A.; Mattei, M.G.; Tisser, A.; Horellou, P.; Benarous;

Nucleic Acids Res. 19, 5117-5123, 1991

A;Title: Identification and expression of the cDNA of KIN17, a zinc-finger gene located

A;Reference number: S18666; MUID:92020193; PMID:1923796

A;Accession: S18666

A;Molecule type: mRNA

A;Residues: 1-391 <ANG>

A;Cross-references: UNIPROT:Q8K339; UNIPARC:UPI000017993P; EMBL:X58472

R;Angulo, J.; Rouer, E.; Benarous, R.; Devoret, R.

Biochimie 73, 251-256, 1991

A;Title: Identification of a mouse cDNA fragment whose expressed polypeptide reacts with

A;Reference number: A43753; MUID:91355299; PMID:1715759

A;Accession: A43753

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 74-273 <ANW>

A;Cross-references: UNIPARC:UPI0000179940

C;Superfamily: KIN17 protein

Query Match 92.6%; Score 1887; DB 2; Length 391;
Best Local Similarity 91.9%; Pred. No. 7.3e-100;
Matches 361; Conservative 15; Mismatches 15; Indels 2; Gaps 1;

Qy	1	MGKSDFLTPKAIANRIKSGKGLRWYCOMQOCQCRDENGPKCHCMSESHORQLLASEN	60
Db	1	MGKSDFLSPKAIANRIKSGKGLQNVRYCOMQOCQCRDENGPKCHCMSESHORQLLASEN	60
Qy	61	PQOFMDYFSEFRNDFLELLRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDFTK	120
Db	61	PQOFMDYFSEFRNDFLELLRRFGTKRVHNNIVYNEYISHREIHNNATQWETLTDFTK	120
Qy	121	WLGREGLCVKDETPKGYIYQIDRDPETIRQLEKKKKODLDEBKTAKFIEQVRRG	180
Db	121	WLGREGLCVKDETPKGYIYQIDRDPETIRQLEKKKKQDLDEBKTAKFIEQVRRG	180
Qy	181	LEGKEQVPFTELSRNDEBKYTNLSKGACSSSGATSSKSSLTGSPALKTICSSASVK	240
Db	181	LEGKEQVPFTELSRNDEBKYTNLSKGACSSSGATSSKSSLTGSPALKTICSSASGK	240
Qy	241	RKSSQSSSTOSKSKKKKKSALDEIMEIEBKKRTARTDYMLQPIIVKIIITKLGKGYHK	300
Db	241	RKSSQSSSAQ--PAKKKKSALDEIMEIEBKKRTARTDNLQPIIVKIIITKLGKGYHK	298
Qy	301	KKAIVKEVIDKYTAVVKVDSGDKLQDTHLETVTPAPGKRILVNLGVNKGVTLES	360
Db	299	KKGVVKEVIDRYTAVVKVMTSDGDLKLDQTHLETVTPAPGKRVLVNLGVNKGVTLES	358

myosin-like protei
centromere protei
troponin T - fruit
membrane protein p
chromosome segrega
hypothetical prote
probable chromosom
h-caldesmon - chic
caldesmon - human
hypothetical prote
IGA-specific metal
probable erythrocy
myosin heavy chain
nuclear histone-bi
hypothetical prote

A:Accession: A84671
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-764 <STO>
A:Cross-references: UNIPROT:Q9XINS; UNIPARC:UPI000017A6B9; GB:AB002093; NID:g5306269; P1
C:Gene: At2g27280
A:Map position: 2

Query Match 6.8%; Score 139.5; DB 2; Length 764;
Best Local Similarity 19.0%; Pred. No. 1.4;
Matches 78; Conservative 77; Mismatches 134; Indels 121; Gaps 17;

QY 46 MSSEHORQLLASNPQF-----MDYSEFPNDLLELRFRFGTK 87
DB 54 IESQHKAL--BEDPSAFSYDEVYDDMKQKAVLPRMQD--REERPRYIQLMKQARR 108
QY 88 RVHNNIVYNEIGHR-----EHIHNNATQMET-----LTDFTKWLGRGLCKVDETPKGW 137
DB 109 EKEHEIVYERKLAKEREKDEHLSFSDKEKFTGYGKRLKEQKWLAEERLRELRE-----163
QY 138 YIQYIDRDPETIRRL-----ELEKKKQDLDDEKTAKEFIEEQVRRGL 181
DB 164 ----ERDDVTKKDLSDFFYNIGKNVAFGAREVEAKEAEKLEBQKAEKLEBQKAEKL 218
QY 182 EGKEQEVPTTFLSRENDEKVTFNLSKGCSSSGATSSKSSSTLGP-----SALKTIGS 235
DB 219 EELRKEY-TRVEKKRSPEKEV-----SPDGEFGSSRSKSLPLEABEQAVSEKMG 270
QY 236 SASVKRESSQSQTOS-----KEKKKKSALDEIME-----YEEKKRT 274
DB 271 DGTERRSSKKEAKEKVPKAINDKREDIAAAKERFLARSTISITNAILMWNKENEQ 330
QY 275 ARTDY-----WLOPEIIVKIITKLGSEK--YHKKKAIVKEVIDKYTA 314
DB 331 RKIPYEMHIVRSLNVLVDQIEWYIGKRFLENVLDPIGKESNFKRDTAKEITDAVTT 390
QY 315 VVKMIDSGDKLDQTHLETVIPAPGRILVNLGNGYRGNE--GTLESINE 362
DB 391 AKERLRETEK-KTESMDVEKVRPS-----TLFPNASFDPSDPLGFLKVPF 435

RESULT 11
E70474
translation initiation factor IF-2 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C:Accession: E70474
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: E70474
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-805 <AQF>
A:Cross-references: UNIPROT:O67825; UNIPARC:UPI0000056795; GB:AB000769; NID:g2984262; P1
A:Experimental source: strain VF5
C:Genetics:
A:Gene: infB
C:Superfamily: translation initiation factor IF-2; translation elongation factor Tu homC
C:Keywords: GTP binding; nucleotide binding; P-loop
F:309-419/Domain: translation elongation factor Tu homology <ETU>
F:315-322/Region: nucleotide-binding motif A (P-loop)
F:416-419/Region: GTP-binding NKXD motif
F:452-454/Region: GTP-binding SAK/L motif
F:321,322,342,416,417,419,452/Binding site: Mg-GTP (Lys, Thr, Asn, Lys, Asp, Ser) #

Query Match 6.8%; Score 138; DB 2; Length 805;
Best Local Similarity 22.4%; Pred. No. 1.8;
Matches 75; Conservative 54; Mismatches 106; Indels 100; Gaps 17;

QY 143 DRDPETIRQLEKKKKQDLDDEKTAKEFIEEQVRRGLGEGKEQEVPTTFLSRENDEK 202
DB 99 EKKPEVIVEIEB-EKKEEKEEKKPKSVBELIKEILEKKEKEKEK-KKVEKERKEEK 156
QY 203 VTFNLSKGCSSSGATSSKSSSTLGSALTKTIGSSASVKR-----KESSQSSTQSKEKK 255
DB 157 VRVVEVKKEEKEEKEEKEEKP-K-IRMSKKREIMRKLHAEVEKEKKOKSKEKEKK 215
QY 256 KKKS-----ALDIMEIEEEK-----KR--TARTDYWLOPEIIVKIITK- 292
DB 216 KKEEVEKIIYDEVITVRLELLOVPANKVIAELMKRGVLATINOPVPEVAVAESEF 275
QY 293 -----KLGEYHKKKKAIIVKEVIDK-----YTAVMKIDSGDKLKLDOHLETV-- 335
DB 276 GYLAEVKKEEELKEEELKEEELQPRPPIVVMGHVDHGKTTLLDRIRKTNVAE 335
QY 336 -----TPAP-GKRILVLN-CGY-----RG-----NEG--- 355
DB 336 REKGGITQIHGASQVELPDGRKITFLDTPGHEAFTTLRARGAKVTDISVLVWAADGVMP 395
QY 356 -TLESINE-KTFSATIVI-----ETGLPKGRR 380
DB 396 QTIEAINHAKAFNPVPIIVAVNKIDKNADPMKVR 430

RESULT 12
H96770
protein heat shock protein P1017.8 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H96770
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federapfel, N.A.; Kaul, S.; White, O.; Alonso;
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H96770
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-630 <STO>
A:Cross-references: UNIPROT:Q9C911; UNIPARC:UPI000000AA4E1; GB:AE005173; NID:g6939229; P1
C:Genetics:
A:Gene: P1017.8
A:Map position: 1

Query Match 6.8%; Score 133.5; DB 2; Length 630;
Best Local Similarity 21.1%; Pred. No. 2.5;
Matches 74; Conservative 54; Mismatches 133; Indels 89; Gaps 13;

QY 27 YCQMCQKQCDENGFKCHCSESHORQLLASNPQFMDYFSEFPNDLLELRFRFGT 86
DB 309 YCIVCSKKFKSEKQMKHESKKHKEK---AELRESFTDYEBENEEDDGLDLSPEV 365
QY 87 KRVHNNIVYNEIYSHREHIMNATQMETLDTFTKWLGRGLCKVDETPKGWYI-----139
DB 366 EELHEKLOEELNIDNEB-----RDVKKEVVGEADTDDEYFVAEEDMQG 409
QY 140 --QVIDRDPETIRQLEKKKKQ-----DLDEKTAKEFIE-----EQVRR 179
DB 410 SSESEDEDDMTLLKKVSGQKQKNVSKKEDEDETEVEIGDTAEFSEFNDQKSTGR 469
QY 180 GLEKGEQEVPTTFLSRENDEKVTFNLSKGCSSSG-----ATSSKSSSTLGSALK 231
DB 470 NKGAK-----ERNKQAGNMDADTSKVQIPGEGNPDENNNATESASGALADSKD 522
QY 232 TIGSSASVKREKSSQSSTQSKEKK-----KKSALDEI-----MEIEEKK--RTA 275

Db 523 EANSMEYDNKSTGRRRRSKGDKNNQGLNELKSSSEADDTQVYNRDMESQDYKKAPRSK 582

QY 276 RTDYLWLPBIIIVKIITKK-----LGEKYHKKAIVKEVIDKYTAVVK 317

Db 583 KSTRGMK----TKGTTKKNSSNECDRCGEBEFESRTKLHKHLADSGHATVK 628

RESULT 13

T34036

hypothetical protein B0041.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

R:Fulcon, R.; Wohldmann, P.

submitted to the EMBL Data Library, April 1997

A:Description: The sequence of C. elegans cosmid B0041.

A:Reference number: Z21466

A:Accession: T34036

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1359 <FUL>

A:Cross-references: UNIPROT:Q9U7B0; UNIPARC:UPI0000060B63; EMBL:AF000196; PIDN:AAC24256.

A:Experimental source: strain Bristol N2; clone B0041

C:Genetics:

A:Gene: CESP:B0041.7

A:Map position: 1

A:Introns: 12/2; 59/2; 248/2; 582/1; 675/1; 733/3; 966/2; 1044/2; 1310/3

Query Match 6.5%; Score 133; DB 2; Length 1359;

Best Local Similarity 23.7%; Pred. No. 6.4;

Matches 57; Conservative 47; Mismatches 111; Indels 26; Gaps 7;

QY 143 DRDPETIRQLELEKKKKQDLDDDEKTAFTIEQVRRGLEKGEQEVPTTTELSRENDESK 202

Db 131 DSDEEREQKSKSKTKQTSSSESESEERKVKKKKNEKSVKKGRAETSESEDEDE 190

QY 203 VTFNLKGCSSSGATSSKSTLGPALSALTIGSSASVKKESQSQTQSKKKK----- 256

Db 191 KPSKSKGLKKKAKSESESESEDEKVKSKKKKVKVKKGESEDEAPEKKTKTEKKR 250

QY 257 -----KKSALDEIMEIEEKKRTARTDYWLQPBIIIVKIITKKLGEYHKKAIVKEVI-- 309

Db 251 SKTSSSESESEKSDREEKESSP-----KPKKKPLAVKKLLSDESEESDV-EVL PQ 304

QY 310 -DKYTAVVKMIDSGD----KLKLDQTHL-ETVIPAGKRIILVNGVRNGEGTLESINEK 363

Db 305 KKKRGAVTLISDEDEKQKSESEASDVBEKVKSKKAKKQSESESGDSSESGSI-TVNRK 363

QY 364 T 364

Db 364 S 364

RESULT 14

T21861

hypothetical protein F36F2.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T21861

R:Cartage, A.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19479

A:Accession: T21861

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1877 <WIL>

A:Cross-references: UNIPROT:O62235; UNIPARC:UPI000017B98A; EMBL:Z81532; PIDN:CAB04326.1;

A:Experimental source: clone F36F2

C:Genetics:

A:Gene: CESP:F36F2.3

A:Map position: 1

A:Introns: 49/3; 86/2; 112/3; 139/2; 235/3; 284/3; 436/3; 507/2; 566/2; 678/2; 1000/3; 1

Query Match 6.5%; Score 131.5; DB 2; Length 1877;

Best Local Similarity 24.2%; Pred. No. 11;

Matches 72; Conservative 43; Mismatches 113; Indels 69; Gaps 11;

QY 143 DRDPETIRQLELEKKKKQDLDDDEKTAFTIEQVRRGLEKGEQEVPTTTELSRENDESK 202

Db 1020 ERDRERRKDDYEKKSKKESDRONEKEKQREKEVEKEHEKDRKRIIVIVEKSEKPRKS 1079

QY 203 VTFNLKGCSSSGATSSKSTLGPALSALTIGSSASVK-----RKSSQSST---- 249

Db 1080 VHERMQK---ADSTSSSRRTTAPSLERKPVSFVASSKPTTNIRVQYSSSSSTKQOE 1136

QY 250 ---QSKEKKKKKALDEIMEIEEKKRTARTDYWLQPBIIIVKIITKKLGEKY-----HKK 301

Db 1137 DEERSKRDRRKADTDVESIGEKKSSRK-----VPKESVDVVKHKSFKIKFDLLDFDK 1192

QY 302 KAIVKEVID-----KYTAVVKMIDSGDKLKDQTHLETVI-----PAPG--- 340

Db 1193 SAEINLLMDIKFEGNDACETAGLKKASGGSFVLKDHVLEFKINGKVAGKIKTPCEGVVT 1252

QY 341 -----KRLVLNGVYRNGEGTLESINEKTFSATI-----VIETGPLKGRVE 382

Db 1253 FQKGLKPGIVLNKGQ-----VIATVSECTHAIVKDMCATCGKDLREKGRAGORKE 1304

RESULT 15

T28676

rhostry protein - Plasmodium yoelii (fragment)

C:Species: Plasmodium yoelii

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T28676; A45521

R:Singha, K.A.; Keen, J.K.; Ogun, S.A.; Holder, A.A.

Mol. Biochem. Parasitol. 76, 329-332, 1996

A:Title: Comparison of two members of a multigene family coding for high-molecular mass 1

A:Reference number: Z20507; MUID:97077455; PMID:8920022

A:Accession: T28676

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2401 <SIN>

A:Cross-references: UNIPROT:Q26216; UNIPARC:UPI000017B647; EMBL:U36927; NID:g1041784; Pfl

R:Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.

Mol. Biochem. Parasitol. 42, 241-246, 1990

A:Title: Identification of the gene for a Plasmodium yoelii rhostry protein. Multiple col

A:Reference number: A45521; MUID:91101660; PMID:2270106

A:Accession: A45521

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 2260-2401 <KEE>

A:Cross-references: UNIPARC:UPI000017B648; GB:M34281

Query Match 6.4%; Score 129.5; DB 2; Length 2401;

Best Local Similarity 21.5%; Pred. No. 19;

Matches 92; Conservative 66; Mismatches 154; Indels 115; Gaps 16;

QY 56 LASENPQQFMDYFSEE-----FRNDF-----FRNDF-----LKL 79

Db 269 LAKEPPYQTKTKIEKKNIEYNTIKSDPKIYVGDIEQLYNEFVSQVSNIEHNIETFI 328

QY 80 LRRRFGTKRVHNNIYVNE---YISHREIHMNATQWETLTDFTKWLGRP----- 125

Db 329 LTLTKIDNVNYYNIQNMETETVKSHLKNLETNKSLSETILDIKIYIGBITNELNKLTD 388

QY 126 -----GLC-KVDETPK-----GWYQIVDRDPETIRQLEL-----EKKKKQDLDEEKT 169

Db 389 FKNKEKGLSNKIDYAKENVQNLVYKSNILEIKKHYNQINIDNITKEKBAKQNYQDFKEH 448

QY 170 AKFIE-----EQVRRGLEKGEQEVPTTTELSRENDEKVTFNLSKGCSSSGATSSKSTL 225

Db 449 MKTIPNEMKYQKPSIEIKMKDEFLSKVKNKNDKVKYKEKVESEHNKFTLTNK---- 504

QY 226 GPSALKTTIGSSASVKRKESSQSSTQSKKKKKKKSALDEIMEIEEKKRTARTDYWLQPEI 285

Db 505 ----IKTEVSDDEIKYENKFNDSNLNETKKSIEEYQNTLTKK----VDDY-----I 552

Qy	286	IVKIITKLGEXHKKKAIVKEVIDKYTAVVMIDSGDKLKD-----OTHLETVI	336
Db	553	KVCLNTNELITNCHNKQTTUKDKLQNOIKTICKETNSIDKIYTDKFNILTDKKELETKF	612
Qy	337	-----PAPGKRILVNGGYRGNEG-TLESINEKTFSATIVIEGPKLGRVEGIGIYE	387
Db	613	TGLSLNNHESNNKELLTYFYDLCANLGNKENMLYKQFNE-----KEKAVEDIKKK	663
Qy	388	--DISKL	392
Db	664	NVDINKI	670

Search completed: November 23, 2005, 16:34:30
Job time : 27.0638 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:53:47 ; Search time 111.616 Seconds
(without alignments)
10694.773 Million cell updates/sec

Title: US-09-555-529-27

Perfect score: 21

Sequence: 1 tcaaaagacaaactgttctggc 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_in.*

3: gb_env.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pr.*

9: gb_ro.*

10: gb_ats.*

11: gb_ey.*

12: gb_un.*

13: gb_vi.*

14: gb_ttg.*

15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	21	100.0	21	6 AX003334	AX003334 Sequence
2	21	100.0	1102	6 AX003309	AX003309 Sequence
3	21	100.0	1386	9 BC028860	BC028860 Mus muscu
4	21	100.0	1390	6 AX003331	AX003331 Sequence
5	21	100.0	1414	9 MMKIN17	X58472 Mouse KIN17
6	21	100.0	2710	9 BC058169	BC058169 Mus muscu
c 7	21	100.0	180882	14 AC114615	AC114615 Mus muscu
8	21	100.0	182400	14 AC124010	AC124010 Mus muscu
c 9	21	100.0	185378	9 AL772367	AL772367 Mouse DNA
10	21	100.0	221547	14 AC119716	AC119716 Rattus no
c 11	19.4	92.4	144118	14 AC11995	AC11995 Rattus no
12	19.4	92.4	146360	8 HS1156N12	AL009047 Human DNA
c 13	19.4	92.4	153215	14 HS24AI7	AL035452 Homo sapi
c 14	18.4	87.6	3606	8 BC043228	BC043228 Homo sapi
15	18.4	87.6	24190	15 CNS09SC4	BX672566 Oryza sat
c 16	18.4	87.6	43261	14 AC145382	AC145382 Oryza sat
c 17	18.4	87.6	94252	15 AC137616	AC137616 Oryza sat
c 18	18.4	87.6	101218	15 AC118284	AC118284 Oryza sat

C 19	18.4	87.6	108295	15 AP005503	AP005503 Oryza sat
C 20	18.4	87.6	110000	15 AP008214_071	Continuation (72 o
C 21	18.4	87.6	110000	15 AP008214_072	Continuation (73 o
C 22	18.4	87.6	110000	15 AP008214_111	Continuation (112
C 23	18.4	87.6	110000	15 AP008215_007	Continuation (8 of
C 24	18.4	87.6	110000	15 AP008215_087	Continuation (88 o
C 25	18.4	87.6	110000	15 AP008217_081	Continuation (82 o
C 26	18.4	87.6	110000	15 AP008217_121	Continuation (122
C 27	18.4	87.6	110000	15 AP008217_253	Continuation (254
C 28	18.4	87.6	110000	15 AP008218_121	Continuation (122
C 29	18.4	87.6	110000	15 AP008218_122	Continuation (123
C 30	18.4	87.6	110000	15 AP008218_162	Continuation (163
C 31	18.4	87.6	110000	15 AP008218_203	Continuation (204
C 32	18.4	87.6	110000	15 AP008207_194	Continuation (195
C 33	18.4	87.6	110000	15 AP008207_338	Continuation (339
C 34	18.4	87.6	110000	15 AP008208_107	Continuation (108
C 35	18.4	87.6	110000	15 AP008209_084	Continuation (85 o
C 36	18.4	87.6	110000	15 AP008209_174	Continuation (175
C 37	18.4	87.6	110000	15 AP008210_207	Continuation (208
C 38	18.4	87.6	110000	15 AP008210_224	Continuation (225
C 39	18.4	87.6	110000	15 AP008210_321	Continuation (322
C 40	18.4	87.6	110000	15 AP008210_329	Continuation (330
C 41	18.4	87.6	110000	15 AP008211_022	Continuation (23 o
C 42	18.4	87.6	110000	15 AP008211_094	Continuation (95 o
C 43	18.4	87.6	110000	15 AP008211_095	Continuation (96 o
C 44	18.4	87.6	110000	15 AP008211_111	Continuation (112
C 45	18.4	87.6	110000	15 AP008211_216	Continuation (217

ALIGNMENTS

RESULT 1
AX003334
LOCUS AX003334 21 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 27 from Patent WO9929845.
ACCESSION AX003334
VERSION AX003334.1 GI:9927147

SOURCE Mus sp.

ORGANISM Mus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE 1 Angulo-Mora, J.F. and Mauffrey, P.

Sequences coding for kin17 protein and their applications

Patent: WO 9929845-A 27 JUN-1999;

ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE

(FR)

FEATURES Location/Qualifiers

source 1..21

/organism="Mus sp."

/mol_type="unassigned DNA"

/db_xref="taxon:10095"

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 21;

Best Local Similarity 100.0%; Pred. No. 9.6;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCAAAGACAACTGTTGCTGGC 21

Db 1 TCAAAGACAACTGTTGCTGGC 21

RESULT 2

AX003309

LOCUS AX003309 1102 bp DNA linear PAT 24-AUG-2000

DEFINITION Sequence 2 from Patent WO9929845.

ACCESSION AX003309

VERSION AX003309.1 GI:9927126

KEYWORDS

SOURCE
ORGANISM
Mus sp.
Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
Angulo-Mora, J.F. and Mauffrey, P.
TITLE
Sequences coding for kin17 protein and their applications
JOURNAL
Patent: WO 9929845-A 2 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)

FEATURES
source
1..1102
/organism="Mus sp."
/mol_type="unassigned DNA"
/db_xref="taxon:10095"

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1102;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 TCAAGACACAACTGTTGCTGGC 21
|||||
150 TCAAGACACAACTGTTGCTGGC 170

Db
150 TCAAGACACAACTGTTGCTGGC 170

RESULT 3
LOCUS
BC028860
DEFINITION
Mus musculus antigenic determinant of rec-A protein, mRNA (CDNA
clone MGC:18413 IMAGE:3673830), complete cds.

ACCESSION
BC028860
VERSION
MGC.
KEYWORDS
SOURCE
Mus musculus (house mouse)

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Caeavant, T.L.,
Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullaly, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahay, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932

TITLE
JOURNAL
PUBMED
REFERENCE
2 (bases 1 to 1386)
Strausberg, R.
AUTHORS
Direct Submission
TITLE
Submitted (01-MAY-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT
Contact: MGC help desk
Email: cgabbs-remail.nih.gov

FEATURES
source
1..1386
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="MGC:18413 IMAGE:3673830"
/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_lib="NCI CGAP_Mam6"
/lab_host="DH10B"
/notes="Vector: PCMV-SPORT6"
1..1386
/gene="Kin"
/feature="synonym: Kin17"
/db_xref="GeneID:16588"
/db_xref="MGI:96676"
20..1195
/gene="Kin"
/codon_start=1
/product="antigenic determinant of rec-A protein"
/protein_id="AAH28860.1"
/db_xref="GI:22135638"
/db_xref="GeneID:16588"
/db_xref="MGI:96676"
/translation="MGKSDFLSPKAIANRIKSLQKLRYWCQKQCRDENGPKCH
CMSESHQQLLASENPQDFYSEEPNDFLELLRRFGTKRVHNNIVVNEYISHR
EHNNATQWETLTDFTKWLGRGLCKVDKPKWYIQYIDRDPETIRQLEKKKK
QDLHDEKTAKEFIEBOVRGLEKEQETPFTLSRENEEKVTFNLNKGAGSGAGT
TSKSSSLGPSALKLLGSNASKRKSSQSQAPAKKKKSAIDMELEEEKKRTATD
AWLQSGIVVKKLITKLGKTHKKGVVKEVIDRTAVVAVVMTDSDGRLLDQTHLETVI
PAPGKRVLVNGGYRGNEGTLSEINEKAFSATVIETGLPKGRVVEGIIQYEDISKLA"

QY
1 TCAAGACACAACTGTTGCTGGC 21
|||||
169 TCAAGACACAACTGTTGCTGGC 189

Db
169 TCAAGACACAACTGTTGCTGGC 189

Query Match
100.0%; Score 21; DB 9; Length 1386;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
Query Match 100.0%; Score 21; DB 9; Length 1386;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 TCAAGACACAACTGTTGCTGGC 21
|||||
169 TCAAGACACAACTGTTGCTGGC 189

Db
169 TCAAGACACAACTGTTGCTGGC 189

RESULT 4
LOCUS
AX003331
DEFINITION
Sequence 24 from Patent WO9929845.
ACCESSION
AX003331
VERSION
AX003331.1 GI:9927146
KEYWORDS
Mus sp.
Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.


```

REFERENCE
AUTHORS      1 Angulo-Mora,J.F. and Mauffrey,P.
TITLE        Sequences coding for kin17 protein and their applications
JOURNAL      Patent: WO 9929845-A 24 17-JUN-1999;
              ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
              (FR)
FEATURES
source       1. 1390
              /location/Qualifiers
              /organism="Mus sp."
              /mol_type="unassigned DNA"
              /db_xref="taxon:10095"
ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 1390;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCAAAGACAACTGTTGCTGGC 21
        |||||
Db      150 TCAAAGACAACTGTTGCTGGC 170

RESULT 5
MMKIN17      1414 bp mRNA linear ROD 30-OCT-1995
LOCUS        Mouse KIN17 mRNA for kin17 protein.
DEFINITION   X58472
VERSION      X58472.1 GI:1045209
KEYWORDS     KIN17 gene; kin17 protein; zinc-finger protein.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      1 Angulo,J.F., Rouer,E., Benarous,R. and Devoret,R.
TITLE        Identification of a mouse cDNA fragment whose expressed polypeptide
              reacts with anti-recA antibodies
JOURNAL      Biochimie 73 (2-3), 251-256 (1991)
PUBMED      1715759
REFERENCE
AUTHORS      2 Angulo,J.F., Rouer,E., Mazin,A., Mattei,M.G., Tissier,A.,
              Horellou,P., Benarous,R. and Devoret,R.
TITLE        Identification and expression of the cDNA of KIN17, a zinc-finger
              gene located on mouse chromosome 2, encoding a new DNA-binding
              protein
JOURNAL      Nucleic Acids Res. 19 (19), 5117-5123 (1991)
PUBMED      1923796
REFERENCE
AUTHORS      3 Angulo,J.
TITLE        Direct Submission
              Submitted (25-MAR-1991) J. Angulo, C N R S, GEMC-Lab d'Enzymologie,
              1 Avenue de la Terrasse, Gif-Sur-Yvette, Ile-de-France, FRANCE
              revised by [4] MGP
REMARK       4 (bases 1 to 1414)
REFERENCE
AUTHORS      Angulo,J.
TITLE        Direct Submission
              Submitted (30-OCT-1995) J. Angulo, C N R S, GEMC-Lab d'Enzymologie,
              1 Avenue de la Terrasse, Gif-Sur-Yvette, Ile-de-France, FRANCE
JOURNAL      On Oct 31, 1995 this sequence version replaced gi:52795.
COMMENT      Location/Qualifiers
FEATURES
source       1. 1414
              /organism="Mus musculus"
              /mol_type="mRNA"
              /db_xref="taxon:10090"
              /cell_line="M966"
              /tissue_type="brain"
              /dev_stage="embryo"
              1. 1414
              /gene="KIN17"
              /pseudo
              <1..1382
              /gene="KIN17"

```

```

CDS
/pseudo
25..1200
/gene="KIN17"
/pseudo
/codon_start=1
/product="kin17 protein"
/db_xref="PSEUDO:CAA41386.1"
106..174
/gene="KIN17"
/notes="zinc-finger of kin17"
/pseudo
234..842
/gene="KIN17"
/notes="kin17(200) fragment"
/pseudo
/evidence=experimental
472..496
/gene="KIN17"
/notes="kin17 helix1 motif"
/pseudo
781..792
/gene="KIN17"
/notes="nuclear localization signal 1"
/pseudo
907..924
/gene="KIN17"
/notes="nuclear localization signal 2"
/pseudo
1382
/gene="KIN17"
polyA_site
ORIGIN
Query Match      100.0%; Score 21; DB 9; Length 1414;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCAAAGACAACTGTTGCTGGC 21
        |||||
Db      174 TCAAAGACAACTGTTGCTGGC 194

RESULT 6
BC058169      2710 bp mRNA linear ROD 13-FEB-2004
LOCUS        Mus musculus antigenic determinant of rec-A protein, mRNA (cdna
DEFINITION   Clone IMAGE:5026202), partial cds.
ACCESSION    BC058169
VERSION      BC058169.1 GI:37194909
KEYWORDS
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
AUTHORS      1 (bases 1 to 2710)
              Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
              Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
              Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
              Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Haleh,F.,
              Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
              Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
              Scheetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S.,
              Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
              Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
              McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
              Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
              Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
              Fahey,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Shevchenko,Y.,
              Sanchez,A., Whiting,M., Madan,A., Young,A.C., Green,E.D.,
              Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Schmutz,J., Myers,R.M.,
              Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Skalka,U.,
              Butterfield,Y.S., Krzywinski,M.I., Skalka,U., Smailus,D.E.,
              Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

```

```

TITLE
REFERENCE
PUBMED
JOURNAL
AUTHORS
AUTHORS
TITLE
JOURNAL
JOURNAL
REMARK
COMMENT

Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 2710)
Strausberg, R.
Direct Submission
Submitted (15-SEP-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabps@mail.nih.gov
Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaapi, R.,
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 123 Row: d Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: Genomescan gene
prediction, Similarity but not identity to protein.
Location/Qualifiers
1. .2710
/organism="Mus musculus"
/mol_type="mRNA"
/strain="mix FVB/N, C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5026202"
/tissue_type="Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI CGAP Mam5"
/lab_host="DH105"
/note="Vector: pCMV-SPORT6"

FEATURES
source
Query Match 100.0%; Score 21; DB 9; Length 2710;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAGACAACTGTTGCTGC 21
|||||
Db 198 TCAAGACAACTGTTGCTGC 218

RESULT 7
AC114615/c
LOCUS
DEFINITION
Mus musculus clone RP24-82119, WORKING DRAFT SEQUENCE, 4 ordered
pieces.
AC114615
AC114615.5 GI:44886278
HTG; HTGS_PHASE2; HTGS DRAFT; HTGS_FULLTOP.
KEYWORDS
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 180882)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, Clone RP24-82119
Unpublished
2 (bases 1 to 180882)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Chopel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Coyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kelle, C., LaRocque, K., Lamazares, R.,
Landers, T., Lenoczky, J., Levine, R., Lindblad-Toh, K., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCartney, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 180882)
Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,
Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., DeArelano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kelle, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (09-MAR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2004 this sequence version replaced gi:24182248.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24315
Center clone name: 82_1_19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191018 bases at least Q40
Consensus quality: 191303 bases at least Q30
Consensus quality: 191388 bases at least Q20

TITLE
JOURNAL

REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

```

Insert size: 178000; agarose-fp
Insert size: 191574; sum-of-contigs
Quality coverage: 12.2 in Q20 bases; agarose-fp
Quality coverage: 11.3 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 89963: contig of 89963 bp in length
* 89964 90063: gap of 100 bp
* 90064 94560: contig of 4497 bp in length
* 94561 94660: gap of 100 bp
* 94561 169601: contig of 74941 bp in length
* 169602 169701: gap of 100 bp
* 169602 180882: contig of 11181 bp in length.
* 169702

FEATURES
source
1. 180882
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-82119"
/clone_lib="RPC1-24 Male Mouse BAC"
misc_feature
1. 89963
/notes="assembly_fragment"
clone_end:SP6
vector_side:left
89964..90063
/estimated_length=100
misc_feature
90064..94560
/notes="assembly_fragment"
94561..94660
/estimated_length=100
misc_feature
94661..169601
/notes="assembly_fragment"
169602..169701
/estimated_length=100
misc_feature
169702..180882
/notes="assembly_fragment"
clone_end:T7
vector_side:right"

ORIGIN
Query Match 100.0%; Score 21; DB 14; Length 180882;
Best Local Similarity 100.0%; Pred. No. 6.7; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
|||||
Db 52817 TCAAAGACAACTGTTGCTGGC 52797

RESULT 8
AC124010 182400 bp DNA linear HTG 25-AUG-2002
LOCUS Mus musculus chromosome UNK clone RP23-119N4, WORKING DRAFT
DEFINITION Mus musculus chromosome UNK clone RP23-119N4, WORKING DRAFT
SEQUENCE 6 unordered pieces.
AC124010
AC124010.2 GI:22475963
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 182400)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone

```

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
2 (bases 1 to 182400)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (06-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 182400)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Aug 25, 2002 this sequence version replaced gi:21327639.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BA0119N04
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 183528 bases at least Q40
Consensus quality: 184210 bases at least Q30
Consensus quality: 184756 bases at least Q20
Insert size: 197000; agarose-fp
Quality coverage: 9.53 in Q20 bases; agarose-fp
Quality coverage: 9.61 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 11868: contig of 11868 bp in length
* 11869 11968: gap of unknown length
* 11969 28906: contig of 16938 bp in length
* 28907 29006: gap of unknown length
* 29007 72773: contig of 43767 bp in length
* 72774 72873: gap of unknown length
* 72874 124926: contig of 52053 bp in length
* 124927 125026: gap of unknown length
* 125027 181903: contig of 56876 bp in length
* 181903 182002: gap of unknown length
* 182003 182400: contig of 398 bp in length.

FEATURES
source

1. 182400
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clones="RP23-119N4"
1. 11868
/notes="assembly_name:Contig86"

misc_feature

11869..11968
/estimated_length=unknown
11969..28906
/notes="assembly_name:Contig87"

gap

misc_feature

28907..29006
/estimated_length=unknown
29007..72773
/notes="assembly_name:Contig88"

gap

misc_feature

72774..72873
/estimated_length=unknown

gap

Williams, G., Willson, R., Wlarczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weises, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstein, G., and Gibbs, R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 221547)
 Worley, K.C.
 Direct Submission
 Submitted (01-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 221547)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Nov 19, 2002 this sequence version replaced gi:23611286. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GWLC
 Center clone name: CH230-274116
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 208013 bases at least Q40
 Consensus quality: 211685 bases at least Q30
 Consensus quality: 213907 bases at least Q20
 Estimated insert size: 199182; sum-of-contigs estimation
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 17 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 198201: contig of 198201 bp in length
 * 198202 198301: gap of unknown length
 * 198302 199749: contig of 1448 bp in length
 * 199750 199849: gap of unknown length
 * 199850 201115: contig of 1266 bp in length
 * 201116 201215: gap of unknown length
 * 201216 202338: contig of 1123 bp in length
 * 202339 202438: gap of unknown length
 * 202439 203500: contig of 1062 bp in length
 * 203501 203600: gap of unknown length
 * 203601 204905: contig of 1305 bp in length
 * 204906 205005: gap of unknown length
 * 205006 205193: contig of 1187 bp in length
 * 205194 206292: gap of unknown length
 * 206293 207485: contig of 1093 bp in length

207386 207485: gap of unknown length
 * 207486 208519: contig of 1034 bp in length
 * 208520 209758: gap of unknown length
 * 209759 209858: gap of unknown length
 * 209859 211441: contig of 1583 bp in length
 * 211442 211542: gap of unknown length
 * 211543 212876: contig of 1335 bp in length
 * 212877 212976: gap of unknown length
 * 212977 214256: contig of 1280 bp in length
 * 214257 214356: gap of unknown length
 * 214357 217003: contig of 2647 bp in length
 * 217004 217104: gap of unknown length
 * 217105 218459: contig of 1356 bp in length
 * 218460 218559: gap of unknown length
 * 218560 219788: contig of 1229 bp in length
 * 219789 219888: gap of unknown length
 * 219889 221547: contig of 1659 bp in length.

FEATURES
 Location/Qualifiers
 source
 1..221547
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-274116"
 1..1272
 /note="wgs end extension
 clone end:Sp6"
 /note="clone_boundary
 clone end:Sp6"
 site:
 end sequence:BZ266352"
 192445..193553
 /note="wgs contig"
 195743..196592
 /note="clone_boundary
 clone end:T7"
 site:
 end sequence:BZ266349"
 198202..198301
 /estimated_length=unknown
 199750..199849
 /estimated_length=unknown
 201116..201215
 /estimated_length=unknown
 202339..202438
 /estimated_length=unknown
 203501..203600
 /estimated_length=unknown
 204906..205005
 /estimated_length=unknown
 206193..206292
 /estimated_length=unknown
 207386..207485
 /estimated_length=unknown
 208520..208619
 /estimated_length=unknown
 209759..209858
 /estimated_length=unknown
 211442..211541
 /estimated_length=unknown
 212877..212976
 /estimated_length=unknown
 214257..214356
 /estimated_length=unknown
 217004..217103
 /estimated_length=unknown
 218460..218559
 /estimated_length=unknown
 219789..219888
 /estimated_length=unknown
 Query Match 100.0%; Score 21; DB 14; Length 221547;

Best Local Similarity 100.0%: Pred. No. 6.6;		Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	TCAAAGACAACTGTTGCTGGC 21	
Db	213161	TCAAAGACAACTGTTGCTGGC 213181	
RESULT 11	AC141995	144118 bp DNA linear HTG 21-MAR-2003	
AC141995/c	Rattus norvegicus clone CH230-293H7, *** SEQUENCING IN PROGRESS		
LOCUS	*** 48 unordered pieces.		
DEFINITION	AC141995	HTG; HTGS PHASE1.	
VERSION	AC141995.1	GI:29135466	
KEYWORDS	Rattus norvegicus	(Norway rat)	
SOURCE	Rattus norvegicus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;		
	Sciurognathi; Murioidea; Muridae; Murinae; Rattus;		
	1 (bases 1 to 144118)		
REFERENCE	Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,		
AUTHORS	Allen, C., Allen, H., Alebrooks, S., Amin, A., Anguiano, D.,		
	Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,		
	Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,		
	Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,		
	Bryant, N., Buhay, C., Burrell, K., Burrell, K., Calderon, E.,		
	Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,		
	Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,		
	Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,		
	Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,		
	Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,		
	Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,		
	Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,		
	Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,		
	Fraser, C. M., Gabisi, A., Gantä, R., Garcia, A., Garner, T., Garza, M.,		
	Georgiev, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,		
	Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,		
	Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,		
	Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,		
	Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,		
	Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,		
	Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,		
	Kowitz, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,		
	Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,		
	Lorensu, L., Louleghed, H., Lozado, R. J., Lu, X., Ma, J.,		
	Maheeshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,		
	Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,		
	Mawhinney, S., Mcleod, M., McNeill, T., Meenen, E., Milosavljevic, A.,		
	Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K.,		
	Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D.,		
	Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okwuonu, G.,		
	Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H.,		
	Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A.,		
	Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E.,		
	Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y.,		
	Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A.,		
	Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S.,		
	Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A.,		
	Sisson, I., Sitter, C. D., Smajda, D., Sneed, A., Sodergren, E.,		
	Song, X., Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A.,		
	Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S.,		
	Tingey, A., Trejos, Z., Umani, K., Valas, R., Vera, V., Villaseña, D.,		
	Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J.,		
	Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R.,		
	Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S.,		
	Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X.,		
	Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R.,		
	Holt, R. A., Smith, H. O., Weinstein, G. and Gibbs, R. A.		
	Direct Submission		
	Unpublished		
REFERENCE	2	(bases 1 to 144118)	

Worley, K.C.
Direct Submission
Submitted (21-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GRCN
Center Clone name: CH230-293H7
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 108995 bases at least Q40
Consensus quality: 114645 bases at least Q30
Consensus quality: 119503 bases at least Q20
Estimated insert size: 114401; sum-of-coverage estimation
Quality coverage: 2x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1394: contig of 1394 bp in length
1395
1494: gap of unknown length
1495
2850: contig of 1356 bp in length
2851
2850: gap of unknown length
2851
3372: contig of 1022 bp in length
3372
4072: gap of unknown length
4073
5309: contig of 1237 bp in length
5310
5409: gap of unknown length
5410
6564: contig of 1155 bp in length
6565
6664: gap of unknown length
6665
8449: contig of 1785 bp in length
8450
8550: gap of unknown length
8550
9904
10003: gap of unknown length
10004
11746: contig of 1743 bp in length
11747
11846: gap of unknown length
11847
13518: contig of 1672 bp in length
13519
13618: gap of unknown length
13619
14970: contig of 1352 bp in length
14971
15070: gap of unknown length
15071
16373: contig of 1303 bp in length
16374
16473: gap of unknown length
16474
18359: contig of 1886 bp in length
18360
18459: gap of unknown length
18460
19821: contig of 1362 bp in length
19822
19921: gap of unknown length
19922
21449: contig of 1528 bp in length
21450
21549: gap of unknown length
21550
23269: contig of 1720 bp in length
23270
23369: gap of unknown length
23370
25710: contig of 2341 bp in length
25711
25810: gap of unknown length
25811
27760: contig of 1950 bp in length
27761
27860: gap of unknown length
27861
29815: contig of 1955 bp in length
29816
29915: gap of unknown length
29916
31213: contig of 1298 bp in length
31214
31313: gap of unknown length
31314
33042: contig of 1729 bp in length
33043
33142: gap of unknown length

* 33143 34194: contig of 1052 bp in length
* 34195 34294: gap of unknown length
* 34295 36020: contig of 1726 bp in length
* 36021 36120: gap of unknown length
* 36121 37671: contig of 1551 bp in length
* 37672 37771: gap of unknown length
* 37772 37759: contig of 1988 bp in length
* 37760 39839: gap of unknown length
* 39860 42442: contig of 2583 bp in length
* 42443 42542: gap of unknown length
* 42543 44482: contig of 1940 bp in length
* 44483 44582: gap of unknown length
* 44583 46351: contig of 1769 bp in length
* 46352 46451: gap of unknown length
* 46452 48712: contig of 2261 bp in length
* 48713 48812: gap of unknown length
* 48813 51252: contig of 2440 bp in length
* 51253 51352: gap of unknown length
* 51353 54232: contig of 2880 bp in length
* 54233 54332: gap of unknown length
* 54333 57136: contig of 2804 bp in length
* 57137 57236: gap of unknown length
* 57237 59833: contig of 2597 bp in length
* 59834 59933: gap of unknown length
* 59934 62619: contig of 2686 bp in length
* 62620 62719: gap of unknown length
* 62720 65087: contig of 2368 bp in length
* 65088 65187: gap of unknown length
* 65188 68329: contig of 3142 bp in length
* 68330 68429: gap of unknown length
* 68430 72825: contig of 4396 bp in length
* 72826 72926: gap of unknown length
* 72927 76196: contig of 3271 bp in length
* 76197 76296: gap of unknown length
* 76297 80327: contig of 4031 bp in length
* 80328 80427: gap of unknown length
* 80429 83596: contig of 3169 bp in length
* 83597 83696: gap of unknown length
* 83697 88053: contig of 4357 bp in length
* 88054 88153: gap of unknown length
* 88154 91623: contig of 3470 bp in length
* 91624 91723: gap of unknown length
* 91724 96501: contig of 4778 bp in length
* 96502 96601: gap of unknown length
* 96602 101193: contig of 4592 bp in length
* 101194 101293: gap of unknown length
* 101294 106768: contig of 5475 bp in length
* 106769 106868: gap of unknown length
* 106869 111636: contig of 4768 bp in length
* 111637 111736: gap of unknown length
* 111737 117842: contig of 6106 bp in length
* 117843 117942: gap of unknown length
* 117943 130224: contig of 12282 bp in length
* 130225 130324: gap of unknown length
* 130325 144118: contig of 13794 bp in length.

FEATURES

source
1. 144118
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-293H7"
1395..1494
/estimated_length=unknown
2851..2950

gap
gap

Query Match 92.4%; Score 19.4; DB 14; Length 144118;
Best Local Similarity 95.2%; Pred. No. 48;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCAAAGACAACTGTTGCTGGC 21
| | | | |
Db 128481 TCAAAGACAGCTGTTGCTGGC 128461

RESULT 12
HS1156N12
LOCUS
DEFINITION

Human DNA sequence from clone RP5-1156N12 on chromosome X, complete sequence.
AL009047
AL009047.1 GI:3355872
HTG.
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)

REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: veg@sanger.ac.uk
Clone requests: clonesrequest@sanger.ac.uk

COMMENT

On Jul 30, 1998 this sequence version replaced gi:3334529.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em., EMBL; Swi., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
RP5-1156N12 is from the library RPCI-5 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: veg@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES

Location/Qualifiers
1. 146360
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCP704N121156"
/db_xref="taxon:9606"
/clone="RP5-1156N12"
/clone_lib="RPCI-5"

ORIGIN

Query Match 92.4%; Score 19.4; DB 8; Length 146360;
Best Local Similarity 95.2%; Pred. No. 48;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 TCAAAGACAACTGTTGCTGGC 21
| | | | |
Db 56392 TCAAAGACAACTGATGCTGGC 56412

RESULT 13
HS24A17
LOCUS
DEFINITION

Homo sapiens chromosome X clone RP6-24A17 map q21.1-21.33.
AL035452
AL035452.7 GI:10045108
HTG; HTGS_PHASE2; HTGS_CANCELLED.

SOURCE	Homo sapiens (human)	REFERENCE	1 (bases 1 to 3606)
ORGANISM	Homo sapiens	AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.P., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McSwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Rulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skaleka,U., Smalusz,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
REFERENCE			Mammalian Gene Collection Program Team
AUTHORS			Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
TITLE			Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL			12477932
PUBMED			2 (bases 1 to 3606)
AUTHORS			NIH MGC Project
TITLE			Direct Submission
JOURNAL			Submitted (09-JAN-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Bethesda, MD 20892-2590, USA
REMARK			NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT			Contact: MGC help desk Email: cgapbs-remail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Tohiyuki and Piero Carninci (RIKEN) cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@bcgsc.bc.ca Martin Hirst, Thomas Zeng, Ryan Morin, Michelle Mokea, Johnson Pang, Diana Mah, Jing Wang, Kieth Fichter, Eric Chuah, Allen Delaney, Rob Kirkpatrick, Agnes Baross, Sarah Barber, Mabel Brown-John, Steve S. Chand, William Chow, Ryan Babakaiff, Dave Wong, Corey Matsuo, Jaclyn Beland, Susan Gibson, Luis delRio, Ruth Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao, Kim MacDonald, Mike R. Mayo, Josh Moran, Diana Palmquist, JR Santos, Duane Smalusz, Jeff Stott, Miranda Tsai, George Yang, Jacquie Schein, Asim Siddiqui, Steven Jones, Rob Holt, Marco Marra.
FEATURES	source		Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 76 Row: i Column: 24. Location/Qualifiers 1..3606 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:44316 IMAGE:5295602" /tissue_type="Testis" /clone_lib="NIH MGC_97" /lab_host="DH10B" /note="Vector: pBluescriptR" 1..3606 /gene="ZNF181" /note="synonym: HHZ181" /db_xref="GeneID:339318" /db_xref="MIM:606741" 1297..2820 /codon_start=1 /product="zinc finger protein 181 (HHZ181)" /protein_id="AAH43228.1"
misc_feature	1..153215		
ORIGIN			
Query Match	92.4%	Score 19.4;	DB 14; Length 153215;
Best Local Similarity	95.2%	Pred. No. 48;	
Matches	20;	Conservative 0;	Mismatches 1; Indels 0; Gaps 0;
QY	1	TCAAAGACAACCTGTTCTGGC 21	
Db	56389	TCAAAGACAACCTGATGCTGGC 56409	
RESULT 14			
BC043228/c			
LOCUS			
DEFINITION	Homo sapiens zinc finger protein 181 (HHZ181), mRNA (cDNA clone MGC:44316 IMAGE:5295602), complete cds.		
ACCESSION	BC043228		
VERSION	BC043228.1	GI:28175103	
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.		


```
/db_xref="GI:71297022"
/db_xref="GeneID:339318"
/db_xref="MIM:606741"
/translation="MMEKLSKGMIPDWESRWENKELSTYKONYDSDSPQTVIIKVV
KQSYFBSNKKNLLEYTEKLEGGKQVDHFRPAILTSRESPTADSVYKYNIFRSTPHS
KSTLSFPQKISAEGNHSHKDYDILKKNLPKKSIVIKNEKVGKLLNSKSGAAPSQGS
LTLPTCNREKIYTCSECGKAFQKQSLNHRHRIHTGKPYECRECGKTFSHGSSLTR
HLISHGEKPYKICGKAFSHVSLTNHQSHTGKPYECMNCGKSFGRVSHLIEHL
RIHTQKLYECRICGKAFTHRSSLIHHQKIHTEGKPYECRECGKAFCCSSHLTRHQRI
HTMEKQYECNKKLUKVFSSLSFLVQHSIHTTEKPYECQCKRGSFNQESLNHLRNHI
RLKPYECISGKAFHSRSLLOHRIHTGKPYECICGKGTFCSCSNLTVHQRIHTGE
KPYKQNECGKAFSGKSNLTAHQRVHNGKPNVSVEKPLDYNMHTYCKESYRRTV"
```

ORIGIN

```
Query Match      87.6%; Score 18.4; DB 8; Length 3606;
Best Local Similarity 95.0%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CAAAGACAACTGTTGCTGGC 21
      |||||
Db      391 CAAAGACAACTGTTGCTGGC 372
```

RESULT 15

```
CNS09SC4      24190 bp      DNA      linear      PLN 29-JAN-2004
LOCUS      Oryza sativa chromosome 12, . Partial sequence from BAC B1004E12 of
DEFINITION      chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza
      sativa (rice), complete sequence.
```

ACCESSION BX672566

VERSION BX672566.2 GI:41393044

KEYWORDS HTG.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 24190)

AUTHORS Choisne, N., Orjeda, G., Cattolico, L., Demange, N., Wincker, P.,

Segurens, B., Pelletier, E., Scarpelli, C., Salanoubat, M.,

Weissenbach, J. and Quetier, F.

Oryza sativa chromosome 12 sequencing

Unpublished

2 (bases 1 to 24190)

Genoscope.

Direct Submission

Submitted (29-JAN-2004) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

On Jan 29, 2004 this sequence version replaced gi:37936142.

Center: Genoscope / Centre National de Sequencage

Center code: GS

Web site: http://www.genoscope.cns.fr/

Contact: SeqRef@genoscope.cns.fr

The following sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : B1097F04 (AC-BX649219)
Downstream BAC (overlapping the SP6 end) : OSJNB0119N22
(AC=AL732380) WARNING : This sequence only represents a part of the

indicated clone. ----- Finishing boundaries

FINISHED SEGMENT STARTS AT BASE 1

FINISHED SEGMENT ENDS AT BASE 24190

FEATURES

source

Location/Qualifiers

1. 24190

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/sub_species="japonica"

/db_xref="taxon:39947"

/chromosome="12"

/clone="B1004E12"

ORIGIN

```
Query Match      87.6%; Score 18.4; DB 15; Length 24190;
Best Local Similarity 95.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

QY 1 TCAAAGACAACTGTTGCTGG 20

|||||

Db 149 TCAAAGAGAACTGTTGCTGG 168

Search completed: November 26, 2005, 19:11:11
Job time : 115.616 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 13.4698 Seconds
(without alignments)
10390.557 Million cell updates/sec

Title: US-09-555-529-27

Perfect score: 21

Sequence: 1 tcaagaacaactgttctggtgc 21

Scoring table: IDENTITY_NUC

Gapop 10,0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_21.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002as.*

7: Geneseqn2002bs.*

8: Geneseqn2003as.*

9: Geneseqn2003bs.*

10: Geneseqn2003cs.*

11: Geneseqn2003ds.*

12: Geneseqn2004as.*

13: Geneseqn2004bs.*

14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	2	AAX85571 PCR prime
2	21	100.0	1102	2	AAX85550 CDNA of a
3	21	100.0	1390	2	AAX85570 CDNA of a
4	21	100.0	1458	2	AAX79936 Murine Ki
5	18.4	87.6	10319	4	AAX27632 DNA encod
6	18.4	87.6	10319	10	ADB94435 Novel hum
7	17.8	84.8	21	2	AAX85568 PCR prime
8	17.8	84.8	1002	2	AAX85552 Probe der
9	17.8	84.8	1002	2	AAX85551 CDNA of a
10	17.8	84.8	1296	2	AAX85549 CDNA of a
11	17.8	84.8	12555	4	ABK42300 Genomic s
12	17.8	84.8	12555	4	ABK42300 Human imm
13	17.8	84.8	12555	9	ADB60456 Connectiv
14	17.8	84.8	13810	4	ABL06020 Drosophil
15	16.8	80.0	329	13	ACF85421 Human SIR
16	16.8	80.0	780	9	ADA29641 DNA encod
17	16.8	80.0	2795	14	ADZ61207 Murine Nu
18	16.8	80.0	56423	10	ADC85728 Human GPC
19	16.8	80.0	80331	4	AAC89559 Human his

20	16.8	80.0	110000	13	ABD32911_3	Continuation (4 of
21	16.8	80.0	203020	14	ADZ12576	Adz12576 Human can
22	16.4	78.1	561	10	ADK63809	Adk63809 Disease t
23	16.4	78.1	561	10	ADK61845	Adk61845 Disease t
C 24	16.4	78.1	939	4	AAH31803	Aah31803 Human Olf
C 25	16.4	78.1	951	5	AAH42275	Aah42275 Human CDN
C 26	16.4	78.1	951	6	ABZ43011	Abz43011 Human GPC
C 27	16.4	78.1	951	6	ABK68432	Abk68432 Human DNA
C 28	16.4	78.1	951	6	ABK37561	Abk37561 DNA encod
C 29	16.4	78.1	951	12	ADG83307	Adg83307 Human Olf
C 30	16.4	78.1	971	6	AAH40634	Aah40634 Human G-p
C 31	16.4	78.1	980	6	ABQ88378	Abq88378 Human G-p
C 32	16.4	78.1	980	6	ABQ88379	Abq88379 Human G-p
C 33	16.4	78.1	980	10	ADI04292	Adi04292 Human G-p
C 34	16.4	78.1	980	10	ADI04294	Adi04294 Human G-p
C 35	16.4	78.1	1351	10	ADC85738	Adc85738 Human GPC
C 36	16.4	78.1	1885	3	AAC42432	Aac42432 Arabidops
C 37	16.4	78.1	1959	13	ADT15885	Adt15885 Rice gene
C 38	16.4	78.1	2000	8	ADA72658	Ada72658 Rice gene
C 39	16.4	78.1	110000	14	ABE39174_01	Continuation (2 of
C 40	16.4	78.1	110000	14	ABE39174_02	Continuation (3 of
C 41	16.4	78.1	110000	14	ABE39175_17	Continuation (18 o
C 42	16.4	78.1	110000	14	ABE39175_18	Continuation (19 o
C 43	16.4	78.1	110000	14	ABE42401_15	Continuation (16 o
C 44	16.4	78.1	110000	14	ABE42737_04	Continuation (5 of
C 45	16.4	78.1	110000	14	ABE35723_1	Continuation (2 of

ALIGNMENTS

RESULT:1

AAX85571

ID AAX85571 standard; cDNA; 21 BP.

XX AAX85571;

XX 07-SEP-1999 (first entry)

DE PCR primer used to amplify human kin17 cDNA sequences.

XX Human; kin17 protein; cell proliferation; fertility;

XX Hyperproliferative disease; protein interaction; curved DNA;

XX HIV replication; HIV integration; repair enzyme; PCR primer; ss.

XX Synthetic.

OS Mus sp.

PN FR2772046-A1.

XX 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell proliferation or fertility.

XX Example 1; Page 12; 69pp; French.

CC PCR primers AAX85571-72 were used to amplify 1000 base pairs of cDNA encoding a human kin17 protein. The mammalian kin17 protein is useful for preparing a medicament for controlling cell proliferation or for controlling fertility. The medicaments can also be used to treat hyperproliferative diseases. Fragments between amino acids 55 and 235 (preferably between amino acids 129 and 228) of a mammalian kin17 protein

CC are useful for regulating the interaction between proteins and curved
 CC DNA. The fragment can be used to block replication of HIV or its
 CC integration into the human genome or to target repair enzymes to curved
 CC DNA sites. Expression vectors for kin17 can be used for controlling cell
 CC proliferation

SQ Sequence 21 BP; 6 A; 5 C; 5 G; 5 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 2; Length 21;
 Best Local Similarity 100.0%; Pred. No. 1.7;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAGACAACTGTTGCTGGC 21
 |||||
 Db 1 TCAAGACAACTGTTGCTGGC 21

RESULT 2
 AAX85550
 ID AAX85550 standard; cDNA; 1102 BP.
 XX
 AC AAX85550;
 XX
 DT 07-SEP-1999 (first entry)
 XX
 DE cDNA of a gene coding for a mouse deleted kin17 protein.

XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme; ss.
 XX

OS Mus sp.

XX FR2772046-A1.
 PN
 PD 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.
 XX
 PR 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
 DR WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell
 PT proliferation or fertility.

XX Claim 4; Page 31; 69pp; French.

XX The present sequence encodes a mouse kin17 protein with amino acids 129-
 CC 228 deleted. The mammalian kin17 protein is useful for preparing a
 CC medicament for controlling cell proliferation or for controlling
 CC fertility. The medicaments can also be used to treat hyperproliferative
 CC diseases. Fragments between amino acids 55 and 235 (preferably between
 CC amino acids 129 and 228) of a mammalian kin17 protein are useful for
 CC regulating the interaction between proteins and curved DNA. The fragment
 CC can be used to block replication of HIV or its integration into the human
 CC genome or to target repair enzymes to curved DNA sites. Expression
 CC vectors for kin17 can be used for controlling cell proliferation

XX Sequence 1102 BP; 373 A; 205 C; 272 G; 252 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 2; Length 1102;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAGACAACTGTTGCTGGC 21
 |||||
 Db 150 TCAAGACAACTGTTGCTGGC 170

RESULT 3
 AAX85570
 ID AAX85570 standard; cDNA; 1390 BP.

XX AAX85570;

XX 07-SEP-1999 (first entry)

XX cDNA of a gene coding for the murine kin17 protein.

XX Mouse; kin17 protein; cell proliferation; fertility;
 KW hyperproliferative disease; protein interaction; curved DNA;
 KW HIV replication; HIV integration; repair enzyme; ss.

XX Mus sp.

XX FR2772046-A1.

XX 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell
 PT proliferation or fertility.

XX Claim 21; Page 35-36; 69pp; French.

XX The present sequence encodes a murine kin17 protein. The mammalian kin17
 CC protein is useful for preparing a medicament for controlling cell
 CC proliferation or for controlling fertility. The medicaments can also be
 CC used to treat hyperproliferative diseases. Fragments between amino acids
 CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
 CC kin17 protein are useful for regulating the interaction between proteins
 CC and curved DNA. The fragment can be used to block replication of HIV or
 CC its integration into the human genome or to target repair enzymes to
 CC curved DNA sites. Expression vectors for kin17 can be used for
 CC controlling cell proliferation

XX Sequence 1390 BP; 482 A; 256 C; 353 G; 299 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 2; Length 1390;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAGACAACTGTTGCTGGC 21
 |||||
 Db 150 TCAAGACAACTGTTGCTGGC 170

RESULT 4

AAQ79936
 ID AAQ79936 standard; cDNA; 1458 BP.

XX AAQ79936;

XX 25-MAR-2003 (revised)

DT 06-SEP-1995 (first entry)

XX Murine Kin17 cDNA.

XX chromosomal rearrangement; kin17 protein; SOS DNA repair system; RecA;
 KW genotoxic agent; zinc finger; DNA binding protein; ss.

XX Mus musculus.

XX PH Location/Qualifiers
 FT primer_bind complement(1..21)
 FT /*tag= b
 FT /note= "Oligo L (AAQ79946) binds to complementary strand"
 FT 22..1434
 FT misc_feature /*tag= 1
 FT /label= kin17 cDNA
 FT /note= "nucleotides 1-1414; the genomic DNA contains at
 FT least two introns within this sequence, see Comments"
 FT 32..49
 FT primer_bind /*tag= c
 FT /note= "Oligo C (AAQ79938) binding site"
 FT 46..1221
 FT CDS /*tag= a
 FT /product= "Kin17"
 FT /note= "N's in the sequence denote illegible residues"
 FT complement(67..86)
 FT primer_bind /*tag= d
 FT /note= "Oligo S (AAQ79947) binds to complementary strand"
 FT 274..297
 FT primer_bind /*tag= e
 FT /note= "Oligo D (AAQ79939) binding site"
 FT complement(339..360)
 FT primer_bind /*tag= f
 FT /note= "Oligo K (AAQ79945) binds to complementary strand"
 FT 451..474
 FT primer_bind /*tag= g
 FT /note= "Oligo J (AAQ79944) binding site"
 FT complement(550..567)
 FT primer_bind /*tag= h
 FT /note= "Oligo E (AAQ79940) binds to complementary strand"
 FT 802..825
 FT primer_bind /*tag= i
 FT /note= "Oligo F (AAQ79941) binding site"
 FT complement(839..862)
 FT primer_bind /*tag= j
 FT /note= "Oligo G (AAQ79942) binds to complementary strand"
 FT complement(1435..1458)
 FT primer_bind /*tag= k
 FT /note= "Oligo B (AAQ79937) binds to complementary strand"
 FR2706487-A1.
 23-DEC-1994.
 15-JUN-1993; 93FR-00007171.
 15-JUN-1993; 93FR-00007171.
 (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
 Angulo-Mora JF, Tissier A, Frelat G, Mauffrey P, Guilly M;
 WPI; 1995-039031/06.
 Purified murine kin17 protein prepn. for detecting chromosomal
 rearrangements - also related antibodies, human and murine DNA, primers,
 probes and vectors, used to assess damage caused by genotoxic agents.
 Claim 9; Page 33; 54pp; French.
 The murine Kin17 protein includes a zinc finger domain (see AAR66766),
 recognises single- and double-stranded DNA (partic. regions of secondary
 structure), has apparent mol. wt. 43 kD and is recognised by both anti-
 kin17 antibodies and antibodies against the RecA protein of E.coli. The
 Kin17 protein is involved in DNA repair; it can be used to monitor
 chromosomal rearrangements following exposure to genotoxic agents. The
 kin17 cDNA sequence AAQ79936 consists of a 1414 nucleotide sequence,
 flanked by primer binding sites; the genomic kin17 DNA is claimed in
 which an intron is inserted at position 137 of the 1414 nucleotide cDNA
 and a second intron is located between nucleotides 339-429 of the cDNA.
 (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 1458 BP; 496 A; 272 C; 372 G; 311 T; 0 U; 7 Other;
 Query Match 100.0%; Score 21; DB 2; Length 1458;
 Best Local Similarity 100.0%; Pred. No. 3.6;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TCAAAGACAACTGTGCTGGC 21
 DB 195 TCAAAGACAACTGTGCTGGC 215
 RESULT 5
 AAS27632/c
 ID AAS27632 standard; DNA; 10319 BP.
 XX AC AAS27632;
 XX DT 07-NOV-2001 (first entry)
 XX DE DNA encoding novel signal transduction pathway protein, Seq ID 1292.
 XX KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
 KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
 KW immune system disorder; rheumatoid arthritis; inflammatory condition;
 KW organ transplant rejection; infection; hepatitis C; blood disorder;
 KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
 KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
 KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
 KW reproductive system; gastrointestinal; liver disorder; AIDS; ds;
 KW acquired immune deficiency syndrome.
 XX OS Homo sapiens.
 XX PN WO200154733-A1.
 XX PD 02-AUG-2001.
 XX PF 17-JAN-2001; 2001WO-US001312.
 XX PR 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218220P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 28-SEP-2000; 2000US-0236127P.
PR 28-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.

PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-02559678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465460/50.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders.
XX
XX Claim 1; SEQ ID NO 1292; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative disorders
CC (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g.
CC Alzheimer's disease, Parkinson's disease), chromosomal abnormalities
CC (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g.
CC glomerulonephritis), cardiovascular disorders (e.g. arrhythmia),
CC respiratory disorders, dermatological disorders, in wound healing,
CC epithelial cell proliferation, endocrine disorders (e.g. Addison's
CC disease), reproductive system disorders, gastrointestinal disorder
CC (inflammatory disorders), liver disorders (cirrhosis), as stimulators of
CC B-cell responsiveness to pathogens, activators of T-cells, to induce
CC higher affinity antibodies, and as a means to induce tumour proliferation
CC in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAS26976-
CC AAS27850 represent novel signal transduction pathway protein coding
CC sequences and PCR primers of the invention
XX
SQ Sequence 10319 BP; 3036 A; 1953 C; 2147 G; 3183 T; 0 U; 0 Other;

Query Match 87.6%; Score 18.4; DB 4; Length 10319;
Best Local Similarity 95.0%; Pred. No. 99;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CAAAGACAACTGTGCTGGC 21
 DB 398 CAAAGACAACTGTGCTGGC 379

RESULT 6
 ADB94435/C
 ID ADB94435 standard; DNA; 10319 BP.
 XX AC ADB94435;
 XX DT 04-DEC-2003 (first entry)
 XX DE Novel human protein: DNA #44.
 XX db; gene; human; autoimmune disease; Parkinson's disease; silicosis;
 XX gastrointestinal disease; atherosclerosis; haemophilia; thrombocytopenia;
 KW immunosuppressive agent; adjuvant; enhance immune response;
 KW higher affinity antibody induction;
 KW increased serum immunoglobulin concentration.
 XX Homo sapiens.
 XX US2002168711-A1.
 XX 14-NOV-2002.
 XX 17-JAN-2001; 2001US-00764868.
 XX 31-JAN-2000; 2000US-0179065P.
 XX 04-FEB-2000; 2000US-0180628P.
 XX 28-JUN-2000; 2000US-0214886P.
 XX 07-JUL-2000; 2000US-0216647P.
 XX 07-JUL-2000; 2000US-0216880P.
 XX 11-JUL-2000; 2000US-0217487P.
 XX 11-JUL-2000; 2000US-0218290P.
 XX 14-JUL-2000; 2000US-0217496P.
 XX 26-JUL-2000; 2000US-0220963P.
 XX 26-JUL-2000; 2000US-0220964P.
 XX 14-AUG-2000; 2000US-0224518P.
 XX 14-AUG-2000; 2000US-0224519P.
 XX 14-AUG-2000; 2000US-0225267P.
 XX 14-AUG-2000; 2000US-0225268P.
 XX 14-AUG-2000; 2000US-0225270P.
 XX 14-AUG-2000; 2000US-0225447P.
 XX 14-AUG-2000; 2000US-0225757P.
 XX 14-AUG-2000; 2000US-0225758P.
 XX 22-AUG-2000; 2000US-0226868P.
 XX 30-AUG-2000; 2000US-0228924P.
 XX 01-SEP-2000; 2000US-0229287P.
 XX 01-SEP-2000; 2000US-0229343P.
 XX 01-SEP-2000; 2000US-0229344P.
 XX 01-SEP-2000; 2000US-0229345P.
 XX 05-SEP-2000; 2000US-0229509P.
 XX 05-SEP-2000; 2000US-0229513P.
 XX 08-SEP-2000; 2000US-0231413P.
 XX 21-SEP-2000; 2000US-0234223P.
 XX 21-SEP-2000; 2000US-0234274P.
 XX 25-SEP-2000; 2000US-0234997P.
 XX 27-SEP-2000; 2000US-0235834P.
 XX 29-SEP-2000; 2000US-0236327P.
 XX 29-SEP-2000; 2000US-0236367P.
 XX 29-SEP-2000; 2000US-0236368P.
 XX 29-SEP-2000; 2000US-0236369P.
 XX 29-SEP-2000; 2000US-0236370P.
 XX 02-OCT-2000; 2000US-0236802P.
 XX 02-OCT-2000; 2000US-0237037P.
 XX 02-OCT-2000; 2000US-0237038P.
 XX 02-OCT-2000; 2000US-0237039P.
 XX 13-OCT-2000; 2000US-0237040P.
 XX 20-OCT-2000; 2000US-0239935P.
 XX 20-OCT-2000; 2000US-0240960P.
 XX 20-OCT-2000; 2000US-0241785P.

PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX Rosen CA, Ruben SM, Barash SC;
 PI WPI; 2003-719985/68.
 XX New isolated polypeptide useful for diagnosing and treating
 PT immunosuppressive conditions such as autoimmune disease and Parkinson's
 PT disease.
 XX Disclosure; SEQ ID NO 1292; 345pp; English.
 XX The invention relates to an isolated polypeptide. The polypeptide is
 CC useful for diagnosing a pathological condition or a susceptibility to a
 CC pathological condition in a subject, by determining the presence or
 CC amount of expression of the polypeptide in a biological sample and
 CC diagnosing a pathological condition or a susceptibility to a pathological
 CC condition based on the presence or amount of expression of the
 CC polypeptide. The polypeptide is also useful for identifying a binding
 CC partner to the polypeptide, which involves contacting the polypeptide
 CC with a binding partner and determining whether the binding partner
 CC effects an activity of the polypeptide. The polypeptide or the nucleic
 CC acid encoding the polypeptide is useful for preventing, treating, or
 CC ameliorating a medical condition, which involves administering the
 CC polypeptide or the nucleic acid to a mammalian subject. The nucleic acid
 CC is useful for diagnosing a pathological condition or a susceptibility to
 CC a pathological condition in a subject, which involves determining the
 CC presence or absence of a mutation in the nucleic acid, and diagnosing a
 CC pathological condition or susceptibility to a pathological condition
 CC based on the presence or absence of the mutation. The polypeptide, the
 CC nucleic acid and an antibody to the polypeptide are useful for treating
 CC autoimmune disease, Parkinson's disease, silicosis, gastrointestinal
 CC disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,
 CC the nucleic acid and the antibody are useful as immunosuppressive agents,
 CC as adjuvants to enhance immune responses, and as agents to induce higher
 CC affinity antibodies and increase serum immunoglobulin concentrations. The
 CC present sequence represents DNA encoding a novel human protein. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format direct from USPTO at
 CC seqdata.uspto.gov/sequence.html?DocID=20020168711.
 XX SQ Sequence 10319 BP; 3036 A; 1953 C; 2147 G; 3183 T; 0 U; 0 Other;

Query Match 87.6%; Score 18.4; DB 10; Length 10319;
 Best Local Similarity 95.0%; Pred. No. 99;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 CAAAGACAACTGTGCTGGC 21
 DB 398 CAAAGACAACTGTGCTGGC 379

RESULT 7
 AAX85568
 ID AAX85568 standard; cDNA; 21 BP.
 XX AC AAX85568;
 XX 07-SEP-1999 (first entry)
 XX PCR primer and probe used to detect human kin17 nucleic acids.
 DE Human; kin17 protein; cell proliferation; fertility; probe;
 XX hyperproliferative disease; protein interaction; curved DNA;
 KW

CC are useful for regulating the interaction between proteins and curved
CC DNA. The fragment can be used to block replication of HIV or its
CC integration into the human genome or to target repair enzymes to curved
CC DNA sites. Expression vectors for kin17 can be used for controlling cell
CC proliferation

SQ Sequence 1002 BP; 343 A; 189 C; 230 G; 240 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 2; Length 1002;
Best Local Similarity 90.5%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCAAAGACAACCTGTTGCTGGC 21
Db 207 TCAGAGACAACCTATTGCTGGC 227

RESULT 10
AAX85549
ID AAX85549 standard; cDNA; 1296 BP.

XX AC AAX85549;

XX DT 07-SEP-1999 (first entry)

XX DE cDNA of a gene coding for the human kin17 protein.

XX KW Human; kin17 protein; cell proliferation; fertility;
XX KW hyperproliferative disease; protein interaction; curved DNA;
XX KW HIV replication; HIV integration; repair enzyme; ss.

XX OS Homo sapiens.

XX PN FR2772046-A1.

XX PD 11-JUN-1999.

XX PF 09-DEC-1997; 97FR-00015536.

XX PR 09-DEC-1997; 97FR-00015536.

XX PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX PS WPI; 1999-359999/31.

XX PT New DNA coding for human kin17 protein - useful for controlling cell
XX proliferation or fertility.

XX PS Claim 1; Page 30-31; 69pp; French.

XX CC The present sequence encodes a human kin17 protein. The mammalian kin17
XX protein is useful for preparing a medicament for controlling cell
XX proliferation or for controlling fertility. The medicaments can also be
XX used to treat hyperproliferative diseases. Fragments between amino acids
XX 55 and 235 (preferably between amino acids 129 and 228) of a mammalian
XX kin17 protein are useful for regulating the interaction between proteins
XX and curved DNA. The fragment can be used to block replication of HIV or
XX its integration into the human genome or to target repair enzymes to
XX curved DNA sites. Expression vectors for kin17 can be used for
XX controlling cell proliferation

SQ Sequence 1296 BP; 453 A; 240 C; 310 G; 293 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 2; Length 1296;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCAAAGACAACCTGTTGCTGGC 21
Db 207 TCAGAGACAACCTATTGCTGGC 227

RESULT 11
ABK42300
ID ABK42300 standard; DNA; 12555 BP.

XX AC ABK42300;

XX DT 21-MAY-2002 (first entry)

XX DE Genomic sequence #199 encoding novel human connective tissue polypeptide.
XX KW Human; connective tissue related disorder; cancer; gene therapy;
XX KW cytoskeletal; gene; ds.

XX OS Homo sapiens.

XX PN WO200155343-A1.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US001322.

XX PR 31-JAN-2000; 2000US-0179065P.

XX PR 04-FEB-2000; 2000US-0180628P.

XX PR 24-FEB-2000; 2000US-0184664P.

XX PR 02-MAR-2000; 2000US-0186350P.

XX PR 16-MAR-2000; 2000US-0189874P.

XX PR 17-MAR-2000; 2000US-0190076P.

XX PR 18-APR-2000; 2000US-0198123P.

XX PR 19-MAY-2000; 2000US-0205515P.

XX PR 07-JUN-2000; 2000US-0209467P.

XX PR 28-JUN-2000; 2000US-0214886P.

XX PR 30-JUN-2000; 2000US-0215135P.

XX PR 07-JUL-2000; 2000US-0216647P.

XX PR 07-JUL-2000; 2000US-0216800P.

XX PR 11-JUL-2000; 2000US-0217487P.

XX PR 14-JUL-2000; 2000US-0218290P.

XX PR 26-JUL-2000; 2000US-0220963P.

XX PR 26-JUL-2000; 2000US-0220964P.

XX PR 14-AUG-2000; 2000US-0224518P.

XX PR 14-AUG-2000; 2000US-0224519P.

XX PR 14-AUG-2000; 2000US-0225213P.

XX PR 14-AUG-2000; 2000US-0225214P.

XX PR 14-AUG-2000; 2000US-0225266P.

XX PR 14-AUG-2000; 2000US-0225267P.

XX PR 14-AUG-2000; 2000US-0225268P.

XX PR 14-AUG-2000; 2000US-0225270P.

XX PR 14-AUG-2000; 2000US-0225447P.

XX PR 14-AUG-2000; 2000US-0225757P.

XX PR 14-AUG-2000; 2000US-0225758P.

XX PR 18-AUG-2000; 2000US-0226279P.

XX PR 22-AUG-2000; 2000US-0226681P.

XX PR 22-AUG-2000; 2000US-0226868P.

XX PR 22-AUG-2000; 2000US-0237182P.

XX PR 23-AUG-2000; 2000US-0227009P.

XX PR 30-AUG-2000; 2000US-0228924P.

XX PR 01-SEP-2000; 2000US-0229287P.

XX PR 01-SEP-2000; 2000US-0229343P.

XX PR 01-SEP-2000; 2000US-0229344P.

XX PR 01-SEP-2000; 2000US-0229345P.

XX PR 05-SEP-2000; 2000US-0229509P.

XX PR 05-SEP-2000; 2000US-0229513P.

XX PR 06-SEP-2000; 2000US-0230437P.

XX PR 06-SEP-2000; 2000US-0230438P.

XX PR 08-SEP-2000; 2000US-0231242P.

XX PR 08-SEP-2000; 2000US-0231243P.

XX PR 08-SEP-2000; 2000US-0231244P.

XX PR 08-SEP-2000; 2000US-0231413P.

XX PR 08-SEP-2000; 2000US-0231414P.

XX PR 08-SEP-2000; 2000US-0232080P.

XX PR 08-SEP-2000; 2000US-0232081P.

```
PR 12-SEP-2000; 2000US-02311968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0234999P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-565190/63.
XX
XX Nucleic acid encoding novel connective tissue associated polypeptides,
PT used in diagnosing, preventing, treating or ameliorating a disorder such
PT as cancer or rheumatoid arthritis.
XX
XX Disclosure; SEQ ID NO 1187; 673pp; English.
XX
XX The present invention relates to the isolation of novel human connective
CC tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide
CC (cDNA and genomic) sequences encoding them. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of diseases associated with connective tissue(s), including
CC cancer. The polynucleotide sequences of the invention are also useful in
CC gene therapy. ABK42102-ABK43116 represent genomic sequences encoding the
CC novel human connective tissue related polypeptides. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 12555 BP; 2632 A; 3328 C; 3392 G; 3203 T; 0 U; 0 Other;
Query Match 84.8%; Score 17.8; DB 4; Length 12555;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TCAAAGACAACTGTGTGCTGGC 21
Db 8515 TCAAAGACAGCTTTTGTCTGGC 8535
RESULT 12
AAK78331
ID AAK78331 standard; DNA; 12555 BP.
XX
AC AAK78331;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33143.
DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US0001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
```

PR	16-MAR-2000;	2000US-0189874P.	PR	13-OCT-2000;	2000US-0239935P.
PR	17-MAR-2000;	2000US-0190076P.	PR	13-OCT-2000;	2000US-0239937P.
PR	18-APR-2000;	2000US-0198123P.	PR	20-OCT-2000;	2000US-0240960P.
PR	19-MAY-2000;	2000US-0205515P.	PR	20-OCT-2000;	2000US-0241221P.
PR	07-JUN-2000;	2000US-0209467P.	PR	20-OCT-2000;	2000US-0241785P.
PR	28-JUN-2000;	2000US-0214886P.	PR	20-OCT-2000;	2000US-0241786P.
PR	30-JUN-2000;	2000US-0215135P.	PR	20-OCT-2000;	2000US-0241808P.
PR	07-JUL-2000;	2000US-0216647P.	PR	20-OCT-2000;	2000US-0241809P.
PR	07-JUL-2000;	2000US-0216880P.	PR	20-OCT-2000;	2000US-0241826P.
PR	11-JUL-2000;	2000US-0217487P.	PR	01-NOV-2000;	2000US-0244617P.
PR	11-JUL-2000;	2000US-0217496P.	PR	08-NOV-2000;	2000US-0246474P.
PR	14-JUL-2000;	2000US-0220963P.	PR	08-NOV-2000;	2000US-0246475P.
PR	26-JUL-2000;	2000US-0220964P.	PR	08-NOV-2000;	2000US-0246476P.
PR	14-AUG-2000;	2000US-0224518P.	PR	08-NOV-2000;	2000US-0246477P.
PR	14-AUG-2000;	2000US-0224519P.	PR	08-NOV-2000;	2000US-0246478P.
PR	14-AUG-2000;	2000US-0225213P.	PR	08-NOV-2000;	2000US-0246523P.
PR	14-AUG-2000;	2000US-0225214P.	PR	08-NOV-2000;	2000US-0246524P.
PR	14-AUG-2000;	2000US-0225266P.	PR	08-NOV-2000;	2000US-0246525P.
PR	14-AUG-2000;	2000US-0225267P.	PR	08-NOV-2000;	2000US-0246526P.
PR	14-AUG-2000;	2000US-0225270P.	PR	08-NOV-2000;	2000US-0246527P.
PR	14-AUG-2000;	2000US-0225271P.	PR	08-NOV-2000;	2000US-0246528P.
PR	14-AUG-2000;	2000US-0225272P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225273P.	PR	08-NOV-2000;	2000US-0246609P.
PR	14-AUG-2000;	2000US-0225274P.	PR	08-NOV-2000;	2000US-0246610P.
PR	14-AUG-2000;	2000US-0225275P.	PR	08-NOV-2000;	2000US-0246611P.
PR	14-AUG-2000;	2000US-0225276P.	PR	08-NOV-2000;	2000US-0246613P.
PR	14-AUG-2000;	2000US-0225277P.	PR	17-NOV-2000;	2000US-0249207P.
PR	14-AUG-2000;	2000US-0225278P.	PR	17-NOV-2000;	2000US-0249208P.
PR	18-AUG-2000;	2000US-0226279P.	PR	17-NOV-2000;	2000US-0249209P.
PR	22-AUG-2000;	2000US-0226681P.	PR	17-NOV-2000;	2000US-0249210P.
PR	22-AUG-2000;	2000US-0226682P.	PR	17-NOV-2000;	2000US-0249211P.
PR	22-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249212P.
PR	23-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249213P.
PR	30-AUG-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249214P.
PR	01-SEP-2000;	2000US-0229287P.	PR	17-NOV-2000;	2000US-0249215P.
PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249216P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249217P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249218P.
PR	03-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249244P.
PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249245P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249264P.
PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0231413P.	PR	17-NOV-2000;	2000US-0249300P.
PR	08-SEP-2000;	2000US-0231414P.	PR	01-DEC-2000;	2000US-0250160P.
PR	08-SEP-2000;	2000US-0232080P.	PR	05-DEC-2000;	2000US-0251030P.
PR	08-SEP-2000;	2000US-0232081P.	PR	05-DEC-2000;	2000US-0251988P.
PR	12-SEP-2000;	2000US-0231968P.	PR	05-DEC-2000;	2000US-0251989P.
PR	14-SEP-2000;	2000US-0232337P.	PR	06-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232338P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0232339P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0232400P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0232401P.	PR	08-DEC-2000;	2000US-0251869P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251989P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251990P.
PR	21-SEP-2000;	2000US-0234223P.	PR	11-DEC-2000;	2000US-0254097P.
PR	21-SEP-2000;	2000US-0234274P.	PR	05-JAN-2001;	2001US-02559678P.
PR	25-SEP-2000;	2000US-0234997P.	XX		
PR	25-SEP-2000;	2000US-0234998P.	XX		
PR	26-SEP-2000;	2000US-0235484P.	PA		
PR	27-SEP-2000;	2000US-0235834P.	XX		
PR	27-SEP-2000;	2000US-0235836P.	PI		
PR	29-SEP-2000;	2000US-0236327P.	XX		
PR	29-SEP-2000;	2000US-0236367P.	DR		
PR	29-SEP-2000;	2000US-0236368P.	XX		
PR	29-SEP-2000;	2000US-0236369P.	XX		
PR	29-SEP-2000;	2000US-0236370P.	PT		
PR	29-SEP-2000;	2000US-0236802P.	PT		
PR	02-OCT-2000;	2000US-0236802P.	XX		
PR	02-OCT-2000;	2000US-0237037P.	XX		
PR	02-OCT-2000;	2000US-0237038P.	PS		
PR	02-OCT-2000;	2000US-0237039P.	XX		
PR	02-OCT-2000;	2000US-0237040P.	CC		
PR	02-OCT-2000;	2000US-0237040P.	CC		

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI, 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis. Disclosure; SEQ ID NO 33143; 3071pp + Sequence Listing; English. AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic

CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 12555 BP; 2632 A; 3328 C; 3392 G; 3203 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 4; Length 12555;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TCAAGACAACTGTTGCTGCC 21

Db 8515 TCAAGACAGCTTTGCTGCC 8535

RESULT 13

ADB60456

ID ADB60456 standard; DNA; 12555 BP.

XX ADB60456;

XX 04-DEC-2003 (first entry)

DE Connective tissue related genomic DNA #199.

KW cyostatic; neuroprotective; nootropic; antiparkinsonian; cardiovascular;
KW antiarteriosclerotic; immunosuppressive; antirheumatic; antiarthritic;
KW antiinflammatory; antiallergic; antiasthmatic; dermatological;
KW nephrotropic; virucide; fungicide; antibacterial; antiparasitic;
KW gene therapy; ds; connective tissues disorder; rheumatoid arthritis;
KW systemic lupus erythematosus; scleroderma; Sjogren's syndrome; cancer;
KW cancer metastasis; neoplasia; leukaemia; neurodegenerative disorder;
KW Alzheimer's disease; Parkinson's disease; cardiovascular disease;
KW atherosclerosis; myocarditis; cardiopulmonary bypass complication;
KW autoimmune disease; multiple sclerosis; allergic reaction; asthma;
KW rhinitis; eczema; inflammatory condition; Crohn's disease; nephritis;
KW gastrointestinal disorder; inflammatory bowel disease;
KW organ transplant rejection; immune system disorder; Bruton's disease;
KW X-linked lymphoproliferative syndrome;
KW B-cell lymphoproliferative disorder; HIV; AIDS; infection;
KW chromosome identification; chromosome mapping;
KW connective tissue related polynucleotide; gene; ds.

OS Homo sapiens.

XX US2003054375-A1.

PN 20-MAR-2003.

PD 07-MAR-2002; 2002US-00092154.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184564P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.

20-OCT-2000; 2000US-02411787P.
PR 20-OCT-2000; 2000US-02411808P.
PR 20-OCT-2000; 2000US-02411809P.
PR 20-OCT-2000; 2000US-02411826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254037P.
PR 05-JAN-2001; 2001US-0259678P.
PR 17-JAN-2001; 2001US-00764847.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Barash SC;
XX WPI; 2003-634869/60.
XX P-PSDB; ADB59396.
XX New connective tissue-related polypeptides and polynucleotides, useful
PT for treating, preventing and/or prognosing e.g. disorders of connective
PT tissue, (e.g. rheumatoid arthritis), cancers, cancer metastases and/or
PT neoplasias.
XX Disclosure; SEQ ID NO 1187; 248pp; English.
PS
XX
XX The invention describes an isolated nucleic acid molecule (I), which
CC comprises a sequence that is at least 95 % identical to a connective
CC tissue-related polynucleotide encoding connective tissue antigens (CTA).
CC The polypeptide or polynucleotide is useful for preventing, treating, or

CC ameliorating medical conditions in a mammal. The connective tissue
CC polypeptides, polynucleotides and antibodies are particularly useful for
CC treating, preventing and/or prognosing disorders of connective tissues
CC (e.g. rheumatoid arthritis, discoid and systemic lupus erythematosus,
CC scleroderma, or Sjogren's syndrome), cancers, cancer metastases and/or
CC neoplasias (e.g. leukaemia), neurodegenerative disorders (e.g.
CC Alzheimer's disease, or Parkinson's disease), cardiovascular diseases
CC (e.g. atherosclerosis, myocarditis or cardiopulmonary bypass
CC complications), autoimmune diseases (e.g. systemic lupus erythematosus,
CC rheumatoid arthritis, or multiple sclerosis), allergic reactions (e.g.
Query Match 84.8%; Score 17.8; DB 9; Length 12555;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TCAAGACAACTGTTGCTGCGC 21
DB 8515 TCAAGACAGACTTTTCTGCGC 8535
RESULT 14
ABL06020
ID ABL06020 standard; cDNA; 13810 BP.
XX
AC ABL06020;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 12542.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
DR P-PSDB; ABB61917.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 12542; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 13810 BP; 4122 A; 2676 C; 2815 G; 4197 T; 0 U; 0 Other;
Query Match 84.8%; Score 17.8; DB 4; Length 13810;
Best Local Similarity 90.5%; Pred. No. 2.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1

TCAAGACAACTGTTGCTGGC 21

Db

9865

TTAAGCCAACTGTTGCTGGC 9865

RESULT 15

ACF85421

ID ACF85421 standard; DNA; 329 BP.

XX

AC ACF85421;

DT

XX 02-JUN-2005 (first entry)

DE

XX Human SIRS/sepsis diagnostic marker DNA fragment 4281.

XX

KW Systemic inflammatory response syndrome; SIRS; antibacterial;

KW immunosuppressive; antiinflammatory; diagnosis; sepsis; ds.

XX

OS Homo sapiens.

XX

FN WO2004087949-A2.

PD

XX 14-OCT-2004.

XX

PF 31-MAR-2004; 2004WO-EP003419.

XX

PR 02-APR-2003; 2003DE-01015031.

PR

08-AUG-2003; 2003DE-01036511.

PR

02-SEP-2003; 2003DE-01040395.

XX

PA (SIRS-) SIRS LAB GMBH.

XX

PI Rusewurm S, Reinhart K, Saluz H, Straube E, Zipfel PF, Daigner H;

XX

WPI; 2004-748070/73.

XX

PT In vitro detection of systemic inflammatory response syndrome and related

PT conditions, for e.g. monitoring progression, comprises detecting abnormal

PT expression of disease-related genes.

XX

PS Disclosure; Page; 75pp; German.

XX

CC The invention relates to a novel method for in vitro detection of

CC systemic inflammatory response syndrome (SIRS). The method comprises

CC detecting abnormal expression of disease-related genes, or their

CC associated peptides. The method of the invention demonstrates

CC antibacterial, immunosuppressive and antiinflammatory applications and

CC may be used for early differential diagnosis, monitoring progression,

CC assessing risk, assessing the likely response to treatment and for post

CC mortem diagnosis of systemic inflammatory response syndrome, sepsis and

CC sepsis-like conditions. The recombinant or synthetic nucleic acid

CC sequences of the invention, or derived proteins or peptides, may be

CC useful as calibrants in assays for the specified diseases, for evaluating

CC activity or toxicity in screening for active agents and/or for

CC preparation of agents for treatment or prevention of the specified

CC diseases. The current sequence is that of a human SIRS/sepsis diagnostic

CC marker DNA fragment of the invention. Note: The sequence data for this

CC patent did not form part of the printed specification, but was obtained

CC in electronic format directly from WIPO at ftp.wipo.int/pub/published

CC pct sequences. Furthermore, a number of arbitrary SEQ ID NO.s are

CC disclosed within the specification, however, these have not been taken

CC into account during indexing due to inconsistencies in application and

CC format

XX

XX Sequence 329 BP; 111 A; 54 C; 73 G; 91 T; 0 U; 0 Other;

Query Match

80.0%;

Score 16.8;

DB 13;

Length 329;

Best Local Similarity

90.0%;

Pred. No. 3.5e+02;

Matches 18;

Conservative 0;

Mismatches 2;

Indels 0;

Gaps 0;

QY

2

CAAAGACAACTGTTGCTGGC 21

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:58:53 ; Search time 117.387 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-27

Perfect score: 21
Sequence: 1 tcaagacaaactgtgctggc 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 [seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_hic.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_est7.*
9: gb_gsa1.*
10: gb_gsa2.*
11: gb_gsa3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	355	5	BY038159
2	21	100.0	407	1	AA684460 vm64d01.8
3	21	100.0	465	1	A1180541 uc70c09.1
4	21	100.0	496	2	B8866972 BB866972
5	21	100.0	500	7	CF916748 CF916748
6	21	100.0	501	6	CA560899 CA560899
7	21	100.0	515	1	AA638405 AA638405
8	21	100.0	538	6	CA559106 CA559106
9	21	100.0	573	7	CF914584 CF914584
10	21	100.0	583	2	BF321725 BF321725
11	21	100.0	600	5	BU919249 BU919249
12	21	100.0	625	7	CF914851 CF914851
13	21	100.0	638	4	AK009429 AK009429
14	21	100.0	653	6	CF172363 CF172363
15	21	100.0	653	6	CF172995 CF172995
16	21	100.0	654	2	BE226431 BE226431
17	21	100.0	660	6	CB847652 CB847652
18	21	100.0	663	8	CX227787 CX227787
19	21	100.0	674	7	CF914272 CF914272
20	21	100.0	696	6	CF174012 CF174012
21	21	100.0	740	8	CX763426 CX763426
22	21	100.0	942	5	BQ714508 BQ714508

23	21	100.0	1077	5	BU511663	BU511663	AGENCOURT
24	19.4	92.4	342	8	DN131260	DN131260	1147889 M
25	19.4	92.4	377	8	DN133257	DN133257	997096 MA
26	19.4	92.4	440	6	CB749078	CB749078	AMGNUUC:M
27	19.4	92.4	454	8	DN127913	DN127913	1141950 M
28	19.4	92.4	487	2	BF396990	BF396990	UT-R-BS2-
29	19.4	92.4	512	2	BE014334	BE014334	146106 MA
30	19.4	92.4	544	6	CB288055	CB288055	CMD73_A09
31	19.4	92.4	630	5	BX924377	BX924377	BX924377 BX924377
32	19.4	92.4	635	5	BX925174	BX925174	BX925174 BX925174
33	19.4	92.4	645	7	DN166982	DN166982	999073 MA
34	19.4	92.4	645	8	DN128273	DN128273	1142334 M
35	19.4	92.4	645	8	DN133982	DN133982	998689 MA
36	19.4	92.4	676	2	BI110637	BI110637	602896112
37	19.4	92.4	723	7	CK474145	CK474145	AGENCOURT
38	19.4	92.4	783	7	DN166221	DN166221	997480 MA
39	19.4	92.4	1543	4	AY609991	AY609991	Sus scrof
40	19	90.5	531	7	CO061759	CO061759	est_k_bre
41	18.4	87.6	406	9	CE014061	CE014061	tigr-g88-
42	18.4	87.6	419	10	CL809118	CL809118	OR_CBa002
43	18.4	87.6	488	10	CL736559	CL736559	OR_BBa007
44	18.4	87.6	535	10	CL726793	CL726793	OR_BBa005
45	18.4	87.6	612	9	AQ288077	AQ288077	nbxw0032C

ALIGNMENTS

RESULT 1

BY038159

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BY038159 RIKEN full-length enriched, CRL-1722 L5178Y-R Mus musculus
CDNA clone I730004D10 5', mRNA sequence.
355 bp mRNA linear EST 06-DEC-2002

BY038159 GI:26143602

EST.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Mus.

1 (bases 1 to 355)

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Otsu, N., Saito, R., Suzuki, H., Yamana, A.,
Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A.,
Schonbach, C., Gojibori, T., Baldarelli, R., Hill, D. P., Bult, C.,
Hume, D. A., Quackenbush, J., Schriml, L. M., Kanpin, A., Matsuda, H.,
Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V.,
Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A.,
Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T.,
Gariboldi, M., Glasi, C., Godzik, A., Gough, J., Grimmond, S.,
Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A.,
Kawaji, H., Kawasawa, Y., Kedzierski, R. M., King, B. L., Kongaya, A.,
Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R.,
Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T.,
Numata, K., Okido, T., Pavan, W. J., Perte, G., Pesole, G.,
Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S.,
Ravasi, T., Reed, J. C., Reid, J., Ring, B. Z., Ringwald, M.,
Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K.,
Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M.,
Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y.,
Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I.,
Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P.,
Hayata, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M.,
Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K.,
Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S.,
Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

JOURNAL

PUBMED

COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/ Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site (http://genome.gsc.riken.go.jp) for further details.	
	FEATURES	source
ORIGIN	Location/Qualifiers 1. .355 /organism="Mus musculus" /mol_type="mRNA" /strain="DBA/2" /db_xref="taxon:10090" /clone="I730004D10" /cell_line="CRL-1722 L5178Y-R" /clone_lib="RIKEN full-length enriched, CRL-1722 L5178Y-R"	
	Query Match 100.0%; Score 21; DB 5; Length 355; Best Local Similarity 100.0%; Pred. No. 18; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
RESULT 2	AA684460 407 bp mRNA linear EST 09-DEC-1997	
	LOCUS vm64d01.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone DEFINITION IMAGE:1003009 5' similar to SW:RTS2_YEAST P40962 RTS2 PROTEIN. [1] ; mRNA sequence.	
ACCESSION	AA684460	
	VERSION AA684460.1 GI:2671046	
KEYWORDS	EST.	
	SOURCE Mus musculus (house mouse)	
ORGANISM	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 407)	
	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.	
AUTHORS	The WashU-HHMI Mouse EST Project	
	Unpublished (1996)	
TITLE	Contact: Marra M/Mouse EST Project	
	WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.	
JOURNAL	Unpublished (1996)	
	COMMENT Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.	
FEATURES	Location/Qualifiers 1. .407 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J x DBA/2J F1" /db_xref="taxon:10090" /clone="IMAGE:1003009" /tissue_type="embryo" /dev_stage="2-cell" /lab_host="DH10B" /clone_lib="Knowles Solter mouse 2 cell" /note="Organ: embryo; Vector: pBluescribe (modified); Site: 1: MluI; Site 2: Sall; Cloned unidirectionally from mRNA prepared from 13,500 2-cell stage embryos. Primer: Sall (drt): 5'-CGGTGACCGTCGACCGTGTGTGTTT-3'. CDNAS were cloned into the MluI/Sall sites of a modified pBluescribe vector using commercial linkers (NEB). Average insert size: 1.2 kb."	
	Query Match 100.0%; Score 21; DB 1; Length 407; Best Local Similarity 100.0%; Pred. No. 18; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
ORIGIN	1 TCAAGACAACTGTTGCTGGC 21 91 TCAAGACAACTGTTGCTGGC 111	
	RESULT 3 AI180541 465 bp mRNA linear EST 08-OCT-1998 LOCUS uc70c09.r1 Soares mammary_gland NDMMG Mus musculus cDNA clone DEFINITION IMAGE:1430992 5' similar to SW:RTS2_YEAST P40962 ZINC FINGER PROTEIN RTS2. ; mRNA sequence.	
ACCESSION	AI180541	
	VERSION AI180541.1 GI:3731179	
KEYWORDS	EST.	
	SOURCE Mus musculus (house mouse)	
ORGANISM	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 465)	
	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.	
AUTHORS	The WashU-HHMI Mouse EST Project	
	Unpublished (1996)	
TITLE	Contact: Marra M/Mouse EST Project	
	WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.	


```

VERSION      CF916748.1  GI:38187950
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
             Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1 (bases 1 to 500)
AUTHORS      Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
TITLE        Construction of long-transcript enriched cDNA libraries from
             submicrogram amounts of total RNAs by a universal PCR amplification
             method
JOURNAL      Genome Res. 11 (9), 1553-1558 (2001)
PUBMED       11544199
COMMENT      Contact: Dawood B. Dudekula
             Laboratory of Genetics
             National Institute on Aging/National Institutes of Health
             333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
             Email: cdna@lgsun.grc.nia.nih.gov
             Plate: B0999 row: H column: 08
             Seq primer: M13 Reverse
             High quality sequence stop: 500
             POLYA=No.

FEATURES     Location/Qualifiers
             source
               1..500
               /organism="Mus musculus"
               /mol_type="mRNA"
               /strain="C57BL/6J"
               /db_xref="niaEST:B0999H08-5"
               /db_xref="taxon:10090"
               /clone="NIA:B0999H08 IMAGE:30481819"
               /dev_stage="Unfertilized Egg"
               /lab_host="DH10B"
               /clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long
               1)"
               /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;
             Site 2: NotI; Mouse cDNA project by the Laboratory of
             Genetics, National Institute on Aging (NIA), Intramural
             Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA).
             This is a long-transcript enriched cDNA library (Ref.
             Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total
             RNAs were extracted from a pool of 1488 unfertilized eggs.
             Double-stranded cDNAs were synthesized with an Oligo(dT)
             primer [Invitrogen]:
             5'-pGACTAGTCTAGATCGAGCGCGCCCTTTT-3',
             treated with T4 DNA polymerase, and purified by
             ethanol-precipitation. The cDNAs were ligated to
             Lone-linker LL-Sal4, purified by phenol/chloroform, and
             separated from free linkers by Centricon 100. Then, the
             cDNAs were amplified by long-range high fidelity PCR using
             Ex Taq polymerase (Takara) with a primer Sal4-S. The
             products were purified by phenol/chloroform and Centricon
             100. The cDNAs were digested with SalI and NotI enzymes
             and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
             vector. The DH10B E. coli host was transformed with the
             ligation mixture by the standard chemical method. The
             average insert size is about 2.5 kb. The library was
             constructed by Yulan Piao."

ORIGIN
Query Match      100.0%; Score 21; DB 7; Length 500;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAGACAACTGTTGCTGGC 21
    |||
Db 136 TCAAGACAACTGTTGCTGGC 156

RESULT 6
CA560899      501 bp mRNA linear EST 19-NOV-2002
LOCUS         K0278H05-5N NIA Mouse Unfertilized Egg cDNA Library (Long) Mus
DEFINITION

```

```

musculus cDNA clone NIA:K0278H05 IMAGE:30052216 5', mRNA sequence.
CA560899
CA560899.1  GI:25105554
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 501)
Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Stagg, C.A.,
Martin, P., Alba, K., Tanaka, T. and Ko, M.S.H.
Systematic Analyses of NIA Mouse Unfertilized Egg cDNA Library
(Long)
Unpublished (2001)
Other ESTs: K0278H05-3
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: K0278 row: H column: 05
Seq primer: M13 Reverse
High quality sequence stop: 501
POLYA=No.

FEATURES     Location/Qualifiers
             source
               1..501
               /organism="Mus musculus"
               /mol_type="mRNA"
               /strain="C57BL/6J"
               /db_xref="niaEST:K0278H05-5N"
               /db_xref="taxon:10090"
               /clone="NIA:K0278H05 IMAGE:30052216"
               /tissue_type="Unfertilized Egg"
               /lab_host="DH10B"
               /clone_lib="NIA Mouse Unfertilized Egg cDNA Library
             (Long)"
               /note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
             NotI; Mouse cDNA project by the Laboratory of Genetics,
             National Institute on Aging (NIA), Intramural Research
             Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
             a long-transcript enriched cDNA library (Ref. Genome Res.
             11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
             extracted from a pool of 1488 unfertilized eggs.
             Double-stranded cDNAs were synthesized with an Oligo(dT)
             primer [Invitrogen]:
             5'-pGACTAGTCTAGATCGAGCGCGCCCTTTT-3',
             treated with T4 DNA polymerase, and purified by
             ethanol-precipitation. The cDNAs were ligated to
             Lone-linker LL-Sal4, purified by phenol/chloroform, and
             separated from free linkers by Centricon 100. Then, the
             cDNAs were amplified by long-range high fidelity PCR using
             Ex Taq polymerase (Takara) with a primer Sal4-S. The
             products were purified by phenol/chloroform and Centricon
             100. The cDNAs were digested with SalI and NotI enzymes
             and cloned into SalI/NotI site of pSPORT1 plasmid vector.
             The DH10B E. coli host was transformed with the ligation
             mixture by the standard chemical method. The average
             insert size is about 2.5 kb. The library was constructed
             by Yulan Piao (NIA)."

ORIGIN
Query Match      100.0%; Score 21; DB 6; Length 501;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAGACAACTGTTGCTGGC 21
    |||
Db 146 TCAAGACAACTGTTGCTGGC 166

RESULT 7
AA638405

```

```

LOCUS      AA638405      515 bp      mRNA      linear      EST 22-OCT-1997
DEFINITION v199f02.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone IMAGE:988827 5' similar to SW:RTS2_YEAST_P40962 RTS2 PROTEIN.
            ; mRNA sequence.
ACCESSION  AA638405
VERSION    AA638405.1 GI:2561993
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 515)
AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE      The WashU-HMI Mouse EST Project
JOURNAL    Unpublished (1996)
COMMENT    Contact: Marra M/Mouse EST Project
            WashU-HMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:561107
            High quality sequence stop: 457.
FEATURES   Location/Qualifiers
            source          1..515
                             /organism="Mus musculus"
                             /mol_type="mRNA"
                             /strain="C57BL/6J x DBA/2J F1"
                             /db_xref="taxon:10090"
                             /clone="IMAGE:988827"
                             /tissue_type="blastocyst"
                             /dev_stage="embryo (pre-implantation)"
                             /lab_host="DH10B"
                             /clone_lib="Knowles Solter mouse blastocyst B1"
                             /notes="Organ: embryo; Vector: pSPORT; Site 1: NotI;
                             Site 2: SalI; Cloned unidirectionally from mRNA prepared
                             from 800 blastocysts. Primer: SalI(dT):
                             5'-CGTCGACCGTCGACGTTTTTTTTTTT-3'. cDNAs were
                             cloned into the NotI/SalI sites of a pSPORT vector (Life
                             Technologies). Two different size selections: B1 (larger
                             inserts) and B3."
ORIGIN
            Query Match      100.0%; Score 21; DB 1; Length 515;
            Best Local Similarity 100.0%; Pred. No. 19;
            Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TCAAAGACAACTGTGCTGGC 21
    |||||
Db 147 TCAAAGACAACTGTGCTGGC 167

RESULT 8
CA559106      538 bp      mRNA      linear      EST 19-NOV-2002
LOCUS        K0254C06-5N NIA Mouse Unfertilized Egg cDNA Library (Long) Mus
DEFINITION   musculus cDNA clone NIA:K0254C06 IMAGE:30049853 5', mRNA sequence.
ACCESSION    CA559106
VERSION      CA559106.1 GI:25103669
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.

LOCUS      AA638405      515 bp      mRNA      linear      EST 22-OCT-1997
DEFINITION v199f02.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
            clone IMAGE:988827 5' similar to SW:RTS2_YEAST_P40962 RTS2 PROTEIN.
            ; mRNA sequence.
ACCESSION  AA638405
VERSION    AA638405.1 GI:2561993
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muridae; Murinae; Mus.

```

```

REFERENCE  1 (bases 1 to 538)
AUTHORS    Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Luo,A., Stagg,C.A.,
            Martin,P., Aiba,K., Tanaka,T. and Ko,M.S.H.
TITLE      Systematic Analyses of NIA Mouse Unfertilized Egg cDNA Library
            (Long)
JOURNAL    Unpublished (2001)
COMMENT    Other ESTs: K0254C06-3
            Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Caswell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@igsun.grc.nia.nih.gov
            Plate: K0254 row: C column: 06
            Seq primer: M13 Reverse
            High quality sequence stop: 538
            POLYA=No.
FEATURES   Location/Qualifiers
            source          1..538
                             /organism="Mus musculus"
                             /mol_type="mRNA"
                             /strain="C57BL/6J"
                             /db_xref="niaEST:K0254C06-5N"
                             /db_xref="taxon:10090"
                             /clone="NIA:K0254C06 IMAGE:30049853"
                             /tissue_type="Unfertilized Egg"
                             /lab_host="DH10B"
                             /clone_lib="NIA Mouse Unfertilized Egg cDNA Library
                             (Long)"
                             /notes="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2:
                             NotI; Mouse cDNA project by the Laboratory of Genetics,
                             National Institute on Aging (NIA), Intramural Research
                             Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is
                             a long-transcript enriched cDNA library (Ref. Genome Res.
                             11: 1553-1558 (2001)). [PMID: 11544199]. Total RNAs were
                             extracted from a pool of 1488 unfertilized eggs.
                             Double-stranded cDNAs were synthesized with an Oligo(dT)
                             primer (Invitrogen):
                             5'-pGACTAGTCTAGTCGCGAGCGCGCCCTTTTTTTTTTTT-3',
                             treated with T4 DNA polymerase, and purified by
                             ethanol-precipitation. The cDNAs were ligated to
                             Loner-linker LL-Sal4, purified by phenol/chloroform, and
                             separated from free linkers by Centricon 100. Then, the
                             cDNAs were amplified by long-range high fidelity PCR using
                             Ex Tag polymerase (Takara) with a primer Sal4-S. The
                             products were purified by phenol/chloroform and Centricon
                             100. The cDNAs were digested with SalI and NotI enzymes
                             and cloned into SalI/NotI site of pSPORT1 plasmid vector.
                             The DH10B E. coli host was transformed with the ligation
                             mixture by the standard chemical method. The average
                             insert size is about 2.5 kb. The library was constructed
                             by Yulan Piao (NIA)."
ORIGIN
            Query Match      100.0%; Score 21; DB 6; Length 538;
            Best Local Similarity 100.0%; Pred. No. 19;
            Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TCAAAGACAACTGTGCTGGC 21
    |||||
Db 146 TCAAAGACAACTGTGCTGGC 166

RESULT 9
CF914584      573 bp      mRNA      linear      EST 05-NOV-2003
LOCUS        B0966B12-5 NIA Mouse Unfertilized Egg cDNA Library (Long) 1) Mus
DEFINITION   musculus cDNA clone NIA:B0966B12 IMAGE:30478583 5', mRNA sequence.
ACCESSION    CF914584
VERSION      CF914584.1 GI:38185786
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```


Tel: 713 792 3646
Fax: 713 790 0329

FEATURES

source
1. .600
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/clone_lib="Mouse E14.5 retina lambda ZAP II Library"

ORIGIN

Query Match 100.0%; Score 21; DB 5; Length 600;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGCGC 21
|||||
Db 188 TCAAAGACAACTGTTGCTGCGC 208

RESULT 12

CF914851
LOCUS B0970B07-5 NIA Mouse Unfertilized Egg cDNA Library (Long 1) Mus
DEFINITION musculus cDNA clone NIA:B0970B07 IMAGE:30478962 5', mRNA sequence.

ACCESSION CF914851
VERSION CF914851.1 GI:38186053
KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.

1 (bases 1 to 625)

Piao Y., Ko N.T., Lim M.K. and Ko M.S.H.
Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method

Genome Res. 11 (9), 1553-1558 (2001)
11544199

JOURNAL

PUBMED

COMMENT

Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: B0970 row: B column: 07
Seq primer: M13 Reverse
High quality sequence stop: 625

POLYA=No.

FEATURES

source

Location/Qualifiers
1. .625
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:B0970B07-5"
/db_xref="taxon:10090"
/clones="NIA:B0970B07 IMAGE:30478962"
/dev_stage="Unfertilized Egg"
/lab_host="DH10B"
/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long 1)"

/note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI;
Site 2: NotI; Mouse cDNA project by the Laboratory of
Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA).
This is a long-transcript enriched cDNA library (Ref.
Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total
RNAs were extracted from a pool of 1488 unfertilized eggs.
Double-stranded cDNAs were synthesized with an oligo(dT)
primer [Invitrogen:
5'-TGACTGTCCTAGTCGGAGCGGCCCTTTTTTTTTTT-3'],
treated with T4 DNA polymerase, and purified by

ethanol-precipitation. The cDNAs were ligated to
lone-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by the standard chemical method. The
average insert size is about 2.5 kb. The library was
constructed by Yulan Piao."

ORIGIN

Query Match 100.0%; Score 21; DB 7; Length 625;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGCGC 21
|||||
Db 178 TCAAAGACAACTGTTGCTGCGC 198

RESULT 13

AK009429

LOCUS

DEFINITION

AK009429 638 bp mRNA linear HTC 03-APR-2004
Mus musculus adult male tongue cDNA, RIKEN full-length enriched
library, clone:2310020017 product:antigenic determinant of rec-A
protein, full insert sequence.

ACCESSION AK009429
VERSION AK009429.1 GI:12844218

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)
Mus musculus
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

1

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

AUTHORS	Adachi, J., Aizawa, K., Akahira, S., Akimura, S., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
	Direct Submission
	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.jp, URL: http://genome.gsc.riken.jp/, tel: 81-45-503-9222, Fax: 81-45-503-9216)
	Please visit our web site (http://genome.gsc.riken.jp/) for further details.
TITLE	CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.
	Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adaptor of sequence [5' GAGAGAGAGATTCGAGTCTTAATAATATCCCTCCCTCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.
JOURNAL	Location/Qualifiers
	1..638
COMMENT	/organism="Mus musculus"
	/mol_type="mRNA"
FEATURES	/strains="C57BL/6J"
	/db_xref="taxon:10090"
source	/clone="2310020017"
	/sex="male"
CDS	/tissue_type="tongue"
	/clone_lib="RIKEN full-length enriched mouse cDNA library"
ORIGIN	/dev_stage="adult"
	52..>636
Query Match	/note="unnamed protein product; antigenic determinant of rec-A protein (MGD MGI:96676)
	putative"
Best Local Similarity	/codon_start=1
	/protein_id="BAB26281.2"
Matches	/db_xref="GI:26364728"
	/translation="MGKSDFLSPKAIANRIKSLGKQLRWYCOMQCKOCRDENGPKCH CMSEHQQLLASNPQPFMDYFSEFNFDELRRPFGTRVHNNIVNEYISHR EHHNNAQTWETLDTFTWLGRGLCKVDETPKWYIYIDRPETIRRLQLEKKKK QDLDDSEKTAKFIEBQVRRLGKQETPVFTLS"
QY	1 TCAGAAGACAACCTGTTGCTGGC 21
Db	201 TCAGAAGACAACCTGTTGCTGGC 221
RESULT 14	CF172363
	B0906D09-5 NIA Mouse Unfertilized Egg cDNA Library (Long 1) Mus
LOCUS	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
DEFINITION	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
ACCESSION	CF172363
	CF172363.1 GI:33281912
VERSION	EST.
	Mus musculus (house mouse)
KEYWORDS	Mus musculus
	Mus musculus
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
	1 (bases 1 to 653)
REFERENCE	Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
	Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method
AUTHORS	Genome Res. 11 (9), 1553-1558 (2001)
	11544199
TITLE	Contact: Dawood B. Dudekula
	Laboratory of Genetics
JOURNAL	National Institute on Aging/National Institutes of Health
	333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
COMMENT	Email: cdna@igsun.grc.nia.nih.gov
	Plate: B0906 row: D column: 09
FEATURES	Seq primer: M13 Reverse
	High quality sequence stop: 653
source	POLYA-No.
	Location/Qualifiers
CDS	1..653
	/organism="Mus musculus"
ORIGIN	/mol_type="mRNA"
	/strains="C57BL/6J"
Query Match	/db_xref="taxon:10090"
	/clone="B0906D09 IMAGE:30472844"
Best Local Similarity	/dev_stage="Unfertilized Egg"
	/lab_host="DHI08"
Matches	/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long 1)"
	/note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 1488 unfertilized eggs. Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen]: 5'-pGACVAGTCTAGATCGGAGCGCCCTTTTTTTTTT-3', treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DHI08 E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.5 kb. The library was constructed by Yulan Piao."
QY	1 TCAGAAGACAACCTGTTGCTGGC 21
Db	168 TCAGAAGACAACCTGTTGCTGGC 188
RESULT 15	CF172995
	LOCUS
DEFINITION	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
ACCESSION	CF172363
	CF172363.1 GI:33281912
VERSION	EST.
	Mus musculus (house mouse)
KEYWORDS	Mus musculus
	Mus musculus
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
	1 (bases 1 to 653)
REFERENCE	Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
	Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method
AUTHORS	Genome Res. 11 (9), 1553-1558 (2001)
	11544199
TITLE	Contact: Dawood B. Dudekula
	Laboratory of Genetics
JOURNAL	National Institute on Aging/National Institutes of Health
	333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
COMMENT	Email: cdna@igsun.grc.nia.nih.gov
	Plate: B0906 row: D column: 09
FEATURES	Seq primer: M13 Reverse
	High quality sequence stop: 653
source	POLYA-No.
	Location/Qualifiers
CDS	1..653
	/organism="Mus musculus"
ORIGIN	/mol_type="mRNA"
	/strains="C57BL/6J"
Query Match	/db_xref="taxon:10090"
	/clone="B0906D09 IMAGE:30472844"
Best Local Similarity	/dev_stage="Unfertilized Egg"
	/lab_host="DHI08"
Matches	/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long 1)"
	/note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 1488 unfertilized eggs. Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen]: 5'-pGACVAGTCTAGATCGGAGCGCCCTTTTTTTTTT-3', treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DHI08 E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.5 kb. The library was constructed by Yulan Piao."
QY	1 TCAGAAGACAACCTGTTGCTGGC 21
Db	168 TCAGAAGACAACCTGTTGCTGGC 188
RESULT 15	CF172995
	LOCUS
DEFINITION	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
ACCESSION	CF172363
	CF172363.1 GI:33281912
VERSION	EST.
	Mus musculus (house mouse)
KEYWORDS	Mus musculus
	Mus musculus
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
	1 (bases 1 to 653)
REFERENCE	Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
	Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method
AUTHORS	Genome Res. 11 (9), 1553-1558 (2001)
	11544199
TITLE	Contact: Dawood B. Dudekula
	Laboratory of Genetics
JOURNAL	National Institute on Aging/National Institutes of Health
	333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
COMMENT	Email: cdna@igsun.grc.nia.nih.gov
	Plate: B0906 row: D column: 09
FEATURES	Seq primer: M13 Reverse
	High quality sequence stop: 653
source	POLYA-No.
	Location/Qualifiers
CDS	1..653
	/organism="Mus musculus"
ORIGIN	/mol_type="mRNA"
	/strains="C57BL/6J"
Query Match	/db_xref="taxon:10090"
	/clone="B0906D09 IMAGE:30472844"
Best Local Similarity	/dev_stage="Unfertilized Egg"
	/lab_host="DHI08"
Matches	/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long 1)"
	/note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 1488 unfertilized eggs. Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen]: 5'-pGACVAGTCTAGATCGGAGCGCCCTTTTTTTTTT-3', treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DHI08 E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.5 kb. The library was constructed by Yulan Piao."
QY	1 TCAGAAGACAACCTGTTGCTGGC 21
Db	168 TCAGAAGACAACCTGTTGCTGGC 188
RESULT 15	CF172995
	LOCUS
DEFINITION	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
ACCESSION	CF172363
	CF172363.1 GI:33281912
VERSION	EST.
	Mus musculus (house mouse)
KEYWORDS	Mus musculus
	Mus musculus
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
	1 (bases 1 to 653)
REFERENCE	Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
	Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method
AUTHORS	Genome Res. 11 (9), 1553-1558 (2001)
	11544199
TITLE	Contact: Dawood B. Dudekula
	Laboratory of Genetics
JOURNAL	National Institute on Aging/National Institutes of Health
	333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
COMMENT	Email: cdna@igsun.grc.nia.nih.gov
	Plate: B0906 row: D column: 09
FEATURES	Seq primer: M13 Reverse
	High quality sequence stop: 653
source	POLYA-No.
	Location/Qualifiers
CDS	1..653
	/organism="Mus musculus"
ORIGIN	/mol_type="mRNA"
	/strains="C57BL/6J"
Query Match	/db_xref="taxon:10090"
	/clone="B0906D09 IMAGE:30472844"
Best Local Similarity	/dev_stage="Unfertilized Egg"
	/lab_host="DHI08"
Matches	/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long 1)"
	/note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 1488 unfertilized eggs. Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen]: 5'-pGACVAGTCTAGATCGGAGCGCCCTTTTTTTTTT-3', treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DHI08 E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.5 kb. The library was constructed by Yulan Piao."
QY	1 TCAGAAGACAACCTGTTGCTGGC 21
Db	168 TCAGAAGACAACCTGTTGCTGGC 188
RESULT 15	CF172995
	LOCUS
DEFINITION	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
ACCESSION	CF172363
	CF172363.1 GI:33281912
VERSION	EST.
	Mus musculus (house mouse)
KEYWORDS	Mus musculus
	Mus musculus
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
	1 (bases 1 to 653)
REFERENCE	Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
	Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method
AUTHORS	Genome Res. 11 (9), 1553-1558 (2001)
	11544199
TITLE	Contact: Dawood B. Dudekula
	Laboratory of Genetics
JOURNAL	National Institute on Aging/National Institutes of Health
	333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
COMMENT	Email: cdna@igsun.grc.nia.nih.gov
	Plate: B0906 row: D column: 09
FEATURES	Seq primer: M13 Reverse
	High quality sequence stop: 653
source	POLYA-No.
	Location/Qualifiers
CDS	1..653
	/organism="Mus musculus"
ORIGIN	/mol_type="mRNA"
	/strains="C57BL/6J"
Query Match	/db_xref="taxon:10090"
	/clone="B0906D09 IMAGE:30472844"
Best Local Similarity	/dev_stage="Unfertilized Egg"
	/lab_host="DHI08"
Matches	/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long 1)"
	/note="Vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 1488 unfertilized eggs. Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen]: 5'-pGACVAGTCTAGATCGGAGCGCCCTTTTTTTTTT-3', treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker LL-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DHI08 E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.5 kb. The library was constructed by Yulan Piao."
QY	1 TCAGAAGACAACCTGTTGCTGGC 21
Db	168 TCAGAAGACAACCTGTTGCTGGC 188
RESULT 15	CF172995
	LOCUS
DEFINITION	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
	musculus cDNA clone NIA:B0906D09 IMAGE:30472844 5', mRNA sequence.
ACCESSION	CF172363
	CF172363.1 GI:33281912
VERSION	EST.
	Mus musculus (house mouse)
KEYWORDS	Mus musculus
	Mus musculus
SOURCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
	1 (bases 1 to 653)
REFERENCE	Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
	Construction of long-transcript enriched cDNA libraries from submicrogram amounts of total RNAs by a universal PCR amplification method
AUTHORS	Genome Res. 11 (9), 1553-1558 (2001)
	11544199
TITLE	Contact: Dawood B. Dudekula
	Laboratory of Genetics
JOURNAL	National Institute on Aging/National Institutes of Health
	333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
COMMENT	Email: cdna@igsun.grc.nia.nih.gov
	Plate: B0906 row: D column: 09
FEATURES	Seq primer: M13 Reverse
	High quality sequence stop: 653
source	POLYA-No.
	Location/Qualifiers
CDS	1..653
	/organism="Mus musculus"
ORIGIN	/mol_type="mRNA"
	/strains="C57BL/6J"
Query Match	/db_xref="taxon:10090"
	/clone="B0906D09 IMAGE:30472844"
Best Local Similarity	/dev_stage="Unfertilized Egg"
	/lab_host="DHI08"
Matches	/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long 1)"

Job time : 120.387 secs

DEFINITION B0917B04-5 NIA Mouse Unfertilized Egg cDNA Library (Long 1) Mus
musculus cDNA clone NIA:B0917B04 IMAGE:30473871 5', mRNA sequence.
ACCESSION CF172995
VERSION CF172995.1 GI:33282544
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidae; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 653)
AUTHORS Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
TITLE Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
JOURNAL Genome Res. 11 (9), 1553-1558 (2001)
PUBMED 11544199
COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov
Plate: B0917 row: B column: 04
Seq primer: M13 Reverse
High quality sequence stop: 653
POLYA=No.
FEATURES
source
1. .653
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:B0917B04-5"
/db_xref="taxon:10090"
/clone="NIA:B0917B04 IMAGE:30473871"
/dev_stage="Unfertilized Egg"
/lab_host="DH10B"
/clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long
1)"
/note="vector: pCMV-SPORT6 (Invitrogen); Site_1: SalI;
Site_2: NotI; Mouse cDNA project by the Laboratory of
Genetics, National Institute on Aging (NIA), Intramural
Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA).
This is a long-transcript enriched cDNA library (Ref.
Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total
RNAs were extracted from a pool of 1488 unfertilized eggs.
Double-stranded cDNAs were synthesized with an Oligo(dT)
primer [Invitrogen]:
5'-pGACTAGTCTAGATCGAGCGCGCCCTTTT-3',
treated with T4 DNA polymerase, and purified by
ethanol-precipitation. The cDNAs were ligated to
Lone-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pCMV-SPORT6 plasmid
vector. The DH10B E. coli host was transformed with the
ligation mixture by the standard chemical method. The
average insert size is about 2.5 kb. The library was
constructed by Yulan Piao."

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 653;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TCAAAGACAACTGTTGCTGGC 21
Db 168 TCAAAGACAACTGTTGCTGGC 188

Search completed: November 27, 2005, 00:57:58

This Page Blank (uspo)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

QM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:02 ; Search time 4.34891 Seconds
(without alignments)
8583.479 Million cell updates/sec

Title: US-09-555-529-27

Perfect score: 21

Sequence: 1 tcaaaqacaactgttctggc 21

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/1 COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5 COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/H COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq:*
- 7: /cgn2_6/ptodata/1/ina/PP COMB.seq:*
- 8: /cgn2_6/ptodata/1/ina/RE COMB.seq:*
- 9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.8	84.8	601	3	US-09-949-002-8571
2	17.8	84.8	601	3	US-09-949-002-8572
3	17.8	84.8	601	3	US-09-949-002-8573
4	17.8	84.8	601	3	US-09-949-002-8574
5	17.8	84.8	26191	3	US-09-949-002-8574
6	17.8	84.8	39428	3	US-09-949-002-8574
7	16.8	80.0	780	3	US-09-328-352-928
8	16.8	80.0	133559	3	US-09-949-016-15845
9	16.2	77.1	384	3	US-09-902-540-2124
10	16.2	77.1	601	3	US-09-949-016-29076
11	16.2	77.1	601	3	US-09-949-016-69609
12	16.2	77.1	617	3	US-09-902-540-1335
13	16.2	77.1	1946	3	US-08-961-527-339
14	16.2	77.1	1075	3	US-09-949-016-3994
15	16.2	77.1	1854	3	US-09-107-433-1993
16	16.2	77.1	1977	3	US-09-583-110-2082
17	16.2	77.1	1998	3	US-09-543-681A-1706
18	16.2	77.1	17328	3	US-09-949-016-15736
19	16.2	77.1	20537	3	US-09-949-016-12887
20	16.2	77.1	20538	3	US-09-949-016-17052
21	16.2	77.1	34628	3	US-09-949-016-12304
22	16.2	77.1	34779	3	US-09-949-016-13787
23	16.2	77.1	145287	3	US-09-949-016-13530
24	16.2	77.1	145287	3	US-09-949-016-13531

C	25	16	76.2	397	3	US-09-513-999C-32957	Sequence 32957, A
	26	15.8	75.2	519	3	US-08-669-408B-7	Sequence 7, Appl1
	27	15.8	75.2	529	3	US-09-621-976-16620	Sequence 16620, A
	28	15.8	75.2	549	3	US-09-621-976-992	Sequence 992, App
	29	15.8	75.2	601	3	US-09-949-016-13880	Sequence 13880, App
	30	15.8	75.2	601	3	US-09-949-016-140951	Sequence 140951, App
	31	15.8	75.2	687	3	US-09-348-796A-3585	Sequence 3585, App
	32	15.8	75.2	822	3	US-09-792-024-12	Sequence 12, Appl1
	33	15.8	75.2	1429	3	US-09-516-143A-5	Sequence 5, Appl1
	34	15.8	75.2	1429	3	US-09-984-205-5	Sequence 5, Appl1
	35	15.8	75.2	1429	3	US-10-759-277-5	Sequence 5, Appl1
	36	15.8	75.2	1429	3	US-09-930-440C-1	Sequence 1, Appl1
	37	15.8	75.2	1749	3	US-09-348-796A-5536	Sequence 5536, App
	38	15.8	75.2	2536	2	US-09-027-013-2	Sequence 2, Appl1
	39	15.8	75.2	2536	3	US-09-244-233-2	Sequence 2, Appl1
	40	15.8	75.2	5092	3	US-09-620-312D-153	Sequence 153, App
	41	15.8	75.2	31469	3	US-09-949-016-13722	Sequence 13722, A
	42	15.8	75.2	37585	3	US-09-949-016-15847	Sequence 15847, A
	43	15.8	75.2	42571	3	US-09-810-347-3	Sequence 3, Appl1
	44	15.8	75.2	58782	3	US-09-949-016-16851	Sequence 16851, A
	45	15.8	75.2	60785	3	US-09-949-016-12774	Sequence 12774, A

ALIGNMENTS

RESULT 1

US-09-949-002-8571
; Sequence 8571, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8571
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-8571

Query Match 84.8%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TCAAAGACAACACTGTTGCTGGC 21
|||
Db 44 TCAAAGACAACACTGATGTTGGC 64
|||

RESULT 2

US-09-949-002-8572
; Sequence 8572, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0

```
; SEQ ID NO 8572
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-8572

Query Match      84.8%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 245 TCAAAGACAACTGATGTTGGC 265

RESULT 3
US-09-949-002-8573
; Sequence 8573, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8573
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-8573

Query Match      84.8%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 350 TCAAAGACAACTGATGTTGGC 370

RESULT 4
US-09-949-002-8574
; Sequence 8574, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8574
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-8574

Query Match      84.8%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 1 TCAAAGACAACTGATGTTGGC 21

Query Match      84.8%; Score 17.8; DB 3; Length 601;
Best Local Similarity 90.5%; Pred. No. 23;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 7638 TCAAAGACAACTGATGTTGGC 7618

RESULT 7
US-09-328-352-928
; Sequence 928, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
```

```
|||||
Db 397 TCAAAGACAACTGATGTTGGC 417

RESULT 5
US-09-949-002-686/c
; Sequence 686, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 686
; LENGTH: 26191
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-686

Query Match      84.8%; Score 17.8; DB 3; Length 26191;
Best Local Similarity 90.5%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 9629 TCAAAGACAACTGATGTTGGC 9609

RESULT 6
US-09-949-002-794/c
; Sequence 794, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 794
; LENGTH: 39428
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-794

Query Match      84.8%; Score 17.8; DB 3; Length 39428;
Best Local Similarity 90.5%; Pred. No. 52;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
    |||||
Db 7638 TCAAAGACAACTGATGTTGGC 7618

RESULT 7
US-09-328-352-928
; Sequence 928, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
```

; ORGANISM: Human

US-09-949-016-69609

Query Match 77.1%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
||||| ||||| ||||| ||||| |||||
DB 81 TCAAAGCCCAACTGTTGTTGAC 61

RESULT 12

US-09-902-540-1335/c
Sequence 1335, Application US/09902540
Patent No. 6833447

GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 1335
LENGTH: 617
TYPE: DNA
ORGANISM: Myxococcus xanthus

US-09-902-540-1335

Query Match 77.1%; Score 16.2; DB 3; Length 617;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
||||| ||||| ||||| ||||| |||||
DB 147 TCAAACCAACTGTTGCTAGC 127

RESULT 13

US-08-961-527-239/c
Sequence 239, Application US/08961527
Patent No. 6420135

GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 239:
SEQUENCE CHARACTERISTICS:
LENGTH: 946 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-239

Query Match 77.1%; Score 16.2; DB 3; Length 946;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
||||| ||||| ||||| ||||| |||||
DB 617 TCAAAGACATCTGTAGATGGC 597

RESULT 14

US-09-949-016-3994
Sequence 3994, Application US/09949016
Patent No. 6812339
GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3994
LENGTH: 1075
TYPE: DNA
ORGANISM: Human
US-09-949-016-3994

Query Match 77.1%; Score 16.2; DB 3; Length 1075;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGGC 21
||||| ||||| ||||| ||||| |||||
DB 712 TCACAGACATCTGTGCTGCC 732

RESULT 15

US-09-107-433-1993/c
Sequence 433, Application US/09107433
Patent No. 6800744

GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE THERAPEUTICS
NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660

```

;
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
;
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/107.433
;   FILING DATE: 30-Jun-1998
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 60/ 085131
;     FILING DATE: May 12, 1998
;     APPLICATION NUMBER: 60/051553
;     FILING DATE: July 2, 1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Ariniello, Pamela Deneke
;     REGISTRATION NUMBER: 40,489
;     REFERENCE/DOCKET NUMBER: GTC-011
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (781)893-5007
;     TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 1993:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1854 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: circular
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;     ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;     NAME/KEY: misc feature
;     LOCATION: (B) LOCATION 1...1854
;   SEQUENCE DESCRIPTION: SEQ ID NO: 1993:
US-09-107-433-1993      77.1%; Score 16.2; DB 3; Length 1854;
Query Match             85.7%; Pred. No. 1.9e+02;
Best Local Similarity   0; Mismatches 3; Indels 0; Gaps 0;
Matches 18; Conservative 0;

Qy      1 TCAAAGACACTGTGTCTGCG 21
Db      1631 TCAAAGACATCTGTAGATGCG 1611

```

Search completed: November 27, 2005, 01:10:50
Job time : 5.34891 secs

This page Blank (uspio)


```
RESULT 2
US-10-027-632-6241
; Sequence 6241, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6241
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(601)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-6241

Query Match      87.6%; Score 18.4; DB 6; Length 601;
Best Local Similarity 95.0%; Pred. No. 63;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2  CAAAGACAACTGTTGCTGGC 21
Db      109  CAAAGACAACTGTTGCTGGC 128

RESULT 3
US-10-437-963-50742/c
; Sequence 50742, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; File of Invention: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50742
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53199C.1
US-10-437-963-50742

Query Match      87.6%; Score 18.4; DB 6; Length 601;
Best Local Similarity 95.0%; Pred. No. 63;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2  CAAAGACAACTGTTGCTGGC 21
Db      109  CAAAGACAACTGTTGCTGGC 128
```

```
RESULT 4
US-10-437-963-50697/c
; Sequence 50697, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; File of Invention: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50697
; LENGTH: 2250
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53158C.1
US-10-437-963-50697

Query Match      87.6%; Score 18.4; DB 7; Length 2250;
Best Local Similarity 95.0%; Pred. No. 80;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  TCAAAGACAACTGTTGCTGG 20
Db      1154  TCAAAGACAACTGTTGCTGG 1135

RESULT 5
US-10-437-963-50728/c
; Sequence 50728, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; File of Invention: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50728
; LENGTH: 2692
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53186C.1
US-10-437-963-50728

Query Match      87.6%; Score 18.4; DB 7; Length 2692;
```


Best Local Similarity 95.0%; Pred. No. 83;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGG 20
|||||
Db 661 TCAAAGAGAACTGTTGCTGG 642

RESULT 6

US-10-437-963-50685/c
; Sequence 50685, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50685
; LENGTH: 2754
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53147C.1
US-10-437-963-50685

Query Match 87.6%; Score 18.4; DB 7; Length 2754;
Best Local Similarity 95.0%; Pred. No. 83;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGG 20
|||||
Db 1226 TCAAAGAGAACTGTTGCTGG 1207

RESULT 7

US-10-437-963-86964/c
; Sequence 86964, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 86964
; LENGTH: 3735
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85957C.1
US-10-437-963-86964

Query Match 87.6%; Score 18.4; DB 7; Length 3735;
Best Local Similarity 95.0%; Pred. No. 88;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TCAAAGACAACTGTTGCTGG 20
|||||
Db 932 TCAAAGAGAACTGTTGCTGG 913

RESULT 8

US-10-437-963-86972/c
; Sequence 86972, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 86972
; LENGTH: 3831
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85964C.1
US-10-437-963-86972

Query Match 87.6%; Score 18.4; DB 7; Length 3831;
Best Local Similarity 95.0%; Pred. No. 89;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACTGTTGCTGG 20
|||||
Db 1088 TCAAAGAGAACTGTTGCTGG 1069

RESULT 9

US-10-437-963-50771/c
; Sequence 50771, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50771
; LENGTH: 3897
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53224C.1
US-10-437-963-50771

Query Match 87.6%; Score 18.4; DB 7; Length 3897;
Best Local Similarity 95.0%; Pred. No. 89;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACACTGTTGCTGG 20
||||| |||||||
Db 2999 TCAAAGACAACACTGTTGCTGG 2980

RESULT 10
US-10-437-963-86963/c
; Sequence 86963, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 86963
; LENGTH: 3930
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_85956C.1
US-10-437-963-86963

Query Match 87.6%; Score 18.4; DB 7; Length 3930;
Best Local Similarity 95.0%; Pred. No. 89;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACACTGTTGCTGG 20
||||| |||||||
Db 875 TCAAAGACAACACTGTTGCTGG 856

RESULT 11
US-10-437-963-50693/c
; Sequence 50693, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50693
; LENGTH: 4275
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53154C.1
US-10-437-963-50693

Query Match 87.6%; Score 18.4; DB 7; Length 4275;
Best Local Similarity 95.0%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACACTGTTGCTGG 20
||||| |||||||
Db 1157 TCAAAGACAACACTGTTGCTGG 1138

RESULT 12
US-10-437-963-50692/c
; Sequence 50692, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50692
; LENGTH: 4347
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53153C.1
US-10-437-963-50692

Query Match 87.6%; Score 18.4; DB 7; Length 4347;
Best Local Similarity 95.0%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACACTGTTGCTGG 20
||||| |||||||
Db 1226 TCAAAGACAACACTGTTGCTGG 1207

RESULT 13
US-10-437-963-50691/c
; Sequence 50691, Application US/10437963
; Publication NO. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50691
; LENGTH: 4353
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53152C.1
US-10-437-963-50691

Query Match 87.6%; Score 18.4; DB 7; Length 4353;
Best Local Similarity 95.0%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCAAAGACAACACTGTTGCTGG 20

```
Db      1226 TCAAAGAGAACTGTTGCTGG 1207
||||| ||||| ||||| |||||
1181 TCAAAGAGAACTGTTGCTGG 1162

Search completed: November 27, 2005, 02:22:43
Job time : 25.2095 secs

RESULT 14
US-10-437-963-50727/c
; Sequence 50727, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50727
; LENGTH: 4497
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53185C.1
US-10-437-963-50727

Query Match      87.6%; Score 18.4; DB 7; Length 4497;
Best Local Similarity 95.0%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCAAAGAGAACTGTTGCTGG 20
||||| ||||| ||||| |||||
Db      1226 TCAAAGAGAACTGTTGCTGG 1207

RESULT 15
US-10-437-963-50704/c
; Sequence 50704, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 50704
; LENGTH: 4506
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53164C.1
US-10-437-963-50704

Query Match      87.6%; Score 18.4; DB 7; Length 4506;
Best Local Similarity 95.0%; Pred. No. 91;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCAAAGAGAACTGTTGCTGG 20
||||| ||||| ||||| |||||
```

This Page Blank (uspo10)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:41:08 ; Search time 7.41177 Seconds
(without alignments)
421.056 Million cell updates/sec

Title: US-09-555-529-27

Perfect score: 21

Sequence: 1 tcaagacaaactgttctgtgc 21

Scoring table:

IDENTITY INUC

Gapop 10,0 , Gapext 1.0

Searched: 3205263 seqs, 74304013 residues

Total number of hits satisfying chosen parameters: 6410526

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New:*

- 1: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq2.*
- 9: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq3.*
- 10: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	15.2	72.4	903	1	US-10-793-626-1899
C 2	15.2	72.4	929	1	US-10-793-626-1311
C 3	15.2	72.4	2997	1	US-10-793-626-3723
C 4	15.2	72.4	3269	1	US-10-793-626-3692
C 5	15.2	72.4	3314	1	US-10-793-626-3927
C 6	14.8	70.5	19	9	US-11-101-244-1021845
C 7	14.8	70.5	19	8	US-11-083-784-1021845
C 8	14.8	70.5	876	1	US-10-793-626-915
C 9	14.8	70.5	1705	1	US-10-510-386-115
C 10	14.8	70.5	3532	1	US-10-793-626-4153
C 11	14.8	70.5	3657	1	US-10-793-626-3888
C 12	14.8	70.5	4010	1	US-10-793-626-4296
C 13	14.4	68.6	19	8	US-11-101-244-146179
C 14	14.4	68.6	19	8	US-11-101-244-146284
C 15	14.4	68.6	19	8	US-11-101-244-146377
C 16	14.4	68.6	19	9	US-11-083-784-146179
C 17	14.4	68.6	19	9	US-11-083-784-146284
C 18	14.4	68.6	19	9	US-11-083-784-146377
C 19	14.2	67.6	19	8	US-11-101-244-759626
C 20	14.2	67.6	19	8	US-11-101-244-1135786
C 21	14.2	67.6	19	8	US-11-083-784-759626
C 22	14.2	67.6	19	9	US-11-083-784-1135786
C 23	14.2	67.6	447	1	US-10-512-109-18

24	14.2	67.6	839	1	US-10-512-109-4	Sequence 4, Appli
25	14.2	67.6	2754	1	US-10-821-234-163	Sequence 163, App
26	14.2	67.6	40439	1	US-10-993-509-1	Sequence 1, Appli
27	14.2	67.6	48000	7	US-11-159-597-20	Sequence 20, Appl
C 28	14.2	67.6	55763	1	US-10-972-766-1	Sequence 1, Appli
C 29	13.8	65.7	19	8	US-11-101-244-154358	Sequence 154358,
C 30	13.8	65.7	19	8	US-11-101-244-154405	Sequence 154405,
C 31	13.8	65.7	19	8	US-11-101-244-1437747	Sequence 1437747,
C 32	13.8	65.7	19	9	US-11-083-784-154358	Sequence 154358,
C 33	13.8	65.7	19	9	US-11-083-784-154405	Sequence 154405,
C 34	13.8	65.7	19	9	US-11-083-784-1437747	Sequence 1437747,
C 35	13.8	65.7	885	1	US-10-858-730-188	Sequence 188, App
C 36	13.8	65.7	2598	1	US-10-467-962B-32	Sequence 32, Appl
C 37	13.8	65.7	3048	1	US-10-957-569-50	Sequence 50, Appl
C 38	13.8	65.7	17410	7	US-11-051-568-3	Sequence 3, Appli
C 39	13.6	64.8	56	7	US-11-013-533-5	Sequence 5, Appli
C 40	13.6	64.8	1293	7	US-11-074-176-131	Sequence 131, App
C 41	13.6	64.8	1420	1	US-10-512-109-5	Sequence 5, Appli
C 42	13.6	64.8	1722	1	US-10-510-386-113	Sequence 113, App
C 43	13.6	64.8	3132	1	US-10-512-109-22	Sequence 22, Appl
C 44	13.6	64.8	5796	1	US-10-821-234-62	Sequence 62, Appl
C 45	13.6	64.8	7512	1	US-10-647-956A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1

US-10-793-626-1899/c
; Sequence 1899, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1899
; LENGTH: 903
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-1899

Query Match 72.4%; Score 15.2; DB 1; Length 903;
Best Local Similarity 85.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 CAAGACAACTGTTGCTGGC 21

Db 118 CACCGATACTGTTGCTGGC 99

RESULT 2

US-10-793-626-1311
; Sequence 1311, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1

```

; SEQ ID NO 1311
;
; LENGTH: 929
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1311

Query Match 72.4%; Score 15.2; DB 1; Length 929;
Best Local Similarity 85.0%; Pred. No. 51;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TCAAAGACAACCTGTTGCTGG 20
   ||| ||||| |||||
Db 754 TCTAAACAACTGTTACTGG 773

RESULT 3
US-10-793-626-3723/c
; Sequence 3723, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3723
; LENGTH: 2997
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3723

Query Match 72.4%; Score 15.2; DB 1; Length 2997;
Best Local Similarity 85.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TCAAAGACAACCTGTTGCTGG 20
   ||| ||||| |||||
Db 176 TCTAAACAACTGTTACTGG 157

RESULT 4
US-10-793-626-3692
; Sequence 3692, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P03480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3692
; LENGTH: 3269
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3692

```

```
Db      19 CAAAGAAACTGTTGTTG 2
||||| ||||| ||||| ||||| ||
RESULT 7
US-11-083-784-1021845/c
; Sequence 1021845, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR FILING DATE: US/10/714,333
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 1021845
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-1021845
Query Match      70.5%; Score 14.8; DB 9; Length 19;
Best Local Similarity 88.9%; Pred. No. 46;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 CAAAGACAACTGTTGCTG 19
||||| ||||| ||||| ||||| ||
Db      19 CAAAGAAACTGTTGTTG 2

RESULT 8
US-10-793-626-915
; Sequence 915, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 915
; LENGTH: 876
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-915
Query Match      70.5%; Score 14.8; DB 1; Length 876;
Best Local Similarity 88.9%; Pred. No. 81;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 CAAAGACAACTGTTGCTG 19
||||| ||||| ||||| ||||| ||
Db      806 CAAAGACTATTGTTGCTG 823

RESULT 9
US-10-510-386-115/c
; Sequence 115, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 115
; LENGTH: 1705
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (507)..(1202)
US-10-510-386-115
Query Match      70.5%; Score 14.8; DB 1; Length 1705;
Best Local Similarity 88.9%; Pred. No. 89;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      4 AAGACAACTGTTGCTGCG 21
||||| ||||| ||||| ||||| ||
Db      1311 AAGCCAACTGTTGCTGCG 1294

RESULT 10
US-10-793-626-4153
; Sequence 4153, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4153
; LENGTH: 3532
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4153
Query Match      70.5%; Score 14.8; DB 1; Length 3532;
Best Local Similarity 88.9%; Pred. No. 99;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 CAAAGACAACTGTTGCTG 19
||||| ||||| ||||| ||||| ||
Db      2808 CAAAGACTATTGTTGCTG 2825

RESULT 11
US-10-793-626-3888
; Sequence 3888, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
```

```
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3888
; LENGTH: 3657
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3888

Query Match          70.5%; Score 14.8; DB 1; Length 3657;
Best Local Similarity 88.9%; Pred. No. 99;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 CAAAGACAACCTGTTGCTG 19
Db      789 CAAAGACTATTGTTGCTG 806

RESULT 12
US-10-793-626-4296/c
; Sequence 4296, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4296
; LENGTH: 4010
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-4296

Query Match          70.5%; Score 14.8; DB 1; Length 4010;
Best Local Similarity 88.9%; Pred. No. 1e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 CAAAGACAACCTGTTGCTG 19
Db      2242 CAAAGACTATTGTTGCTG 2225

RESULT 13
US-11-101-244-146179/c
; Sequence 146179, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
```

```
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 146179
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-146179

Query Match          68.6%; Score 14.4; DB 8; Length 19;
Best Local Similarity 93.8%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCAAAGACAACCTGTTG 16
Db      18 TCAAAGACAGCTGTTG 3

RESULT 14
US-11-101-244-146284/c
; Sequence 146284, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 146284
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-146284

Query Match          68.6%; Score 14.4; DB 8; Length 19;
Best Local Similarity 93.8%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCAAAGACAACCTGTTG 16
Db      18 TCAAAGACAGCTGTTG 3

RESULT 15
US-11-101-244-146377/c
; Sequence 146377, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
```



```

; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 146377
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-146377

Query Match      68.6%; Score 14.4; DB 8; Length 19;
Best Local Similarity 93.8%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCAAAGACAACTGTTG 16
        |||||
Db      18 TCAAAGACAGCTGTTG 3

```

Search completed: November 27, 2005, 02:44:31
Job time : 7.41177 secs

This Page Blank (uspiid)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:53:47 ; Search time 116.932 Seconds
(without alignments)
10694.773 Million cell updates/sec

Title: US-09-555-529-28

Perfect score: 22

Sequence: 1 atacottcaactctgcgtcctt 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_in.*

3: gb_env.*

4: gb_ov.*

5: gb_pat.*

6: gb_ph.*

7: gb_pr.*

8: gb_ro.*

9: gb_ro.*

10: gb_sts.*

11: gb_sy.*

12: gb_un.*

13: gb_vi.*

14: gb_htg.*

15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	6	AX003335 Sequence
C 2	22	100.0	333	6	AX003341 Sequence
C 3	22	100.0	1102	6	AX003309 Sequence
C 4	22	100.0	1386	9	BC028860 Mus muscu
C 5	22	100.0	1390	6	AX003331 Sequence
C 6	22	100.0	1414	9	MMKIN17
C 7	22	100.0	2710	9	BC058169 Mus muscu
C 8	22	100.0	180882	14	AC114615
C 9	22	100.0	182400	14	AC124010 Mus muscu
C 10	22	100.0	185378	9	AL772367
C 11	22	100.0	221547	14	AC119716
C 12	20.4	92.7	22	6	AX003328 Sequence
C 13	20.4	92.7	333	6	AX003340 Sequence
C 14	20.4	92.7	1002	6	AX003310 Sequence
C 15	20.4	92.7	1002	6	AX003311 Sequence
C 16	20.4	92.7	1528	6	AX003308 Sequence
C 17	20.4	92.7	1528	8	HSBJ5273
C 18	20.4	92.7	2069	8	BC017309 Homo sapi

19	20.4	92.7	174361	8	AL158044
20	20.4	92.7	194563	14	AL391687
C 21	18.8	85.5	675	6	AR547015
C 22	18.8	85.5	3035	13	BYDVPDIP
23	18.8	85.5	5911	15	AF065434
C 24	18.8	85.5	12455	13	AF002227
C 25	18.8	85.5	12271	13	AY149215
C 26	18.8	85.5	12320	13	AY149216
C 27	18.8	85.5	12332	6	AX741975
C 28	18.8	85.5	12332	13	AF502399
C 29	18.8	85.5	12333	13	AF145967
30	18.8	85.5	59538	14	AC012951
31	18.8	85.5	110000	2	AE003524_1
32	18.8	85.5	144118	14	AC141995
C 33	18.8	85.5	164719	2	AC009374
C 34	18.8	85.5	223233	14	AC097562
C 35	18.8	85.5	239787	14	AC160550
36	18.4	83.6	3994	9	BX842664
37	18.4	83.6	123851	9	AL844903
C 38	18.4	83.6	257317	14	AC157305
C 39	18.4	83.6	274906	14	AC094962
C 40	17.8	80.9	10552	13	AY442521
C 41	17.8	80.9	12513	13	BVU18059
C 42	17.8	80.9	38602	8	BX248133
C 43	17.8	80.9	45881	14	AC149956
C 44	17.8	80.9	57997	8	CR925765
45	17.8	80.9	59351	14	AC100656

ALIGNMENTS

RESULT 1
AX003335
LOCUS AX003335 22 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 28 from Patent WO9929845.
ACCESSION AX003335
VERSION AX003335.1 GI:9927148

SOURCE Mus sp.

ORGANISM Mus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 Angulo-Mora, J.F. and Mauffrey, P.

AUTHORS Sequences coding for kin17 protein and their applications

TITLE Patent: WO 9929845-A 28 17-JUN-1999;

JOURNAL ANGULO MORA JAIME FRANCISCO (FR) ; COMMISSARIAT ENERGIE ATOMIQUE

(FR)

FEATURES Location/Qualifiers

source

1..22

/organism="Mus sp."

/mol_type="unassigned DNA"

/db_xref="taxon:10095"

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 22;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22

1 |||||

Db 1 ATACCTTCAACTCTGCGTCCTT 22

1 |||||

RESULT 2

AX003341/c

LOCUS AX003341 333 bp DNA linear PAT 24-AUG-2000

DEFINITION Sequence 34 from Patent WO9929845.

ACCESSION AX003341

VERSION AX003341.1 GI:9927154

KEYWORDS

/protein_id="AAH28860.1"
 /db_xref="GI:22135638"
 /db_xref="GeneID:16588"
 /db_xref="MGI:96676"
 /translation="MGKSDFLSPKAIANRIKSKGLQRLRWYCOMQKQCRDENGFKCH
 CMSHQROLLASNPQDMFYFSEFRNDFLELLRRFGTKRVHNNIVYNYISHR
 EHTHMATQWETLDTFKLGRGLCKVDETPKGMYIQYIDRDPETIRQLELEKKKK
 ODLDDEKTAFTFEORVRLGLEKQETPVFTLSRENEEKVTFNLNKGAGGSAGAT
 TSKSSLSGPSALKLASAGSKRSSQSAQAKKKSALDIMELEBEKRTATD
 AWLPQGVIVKRIITKKLGKHKKGVVKEVIDRYTAVRWMTDSDRLKLDQTHLETVI
 PAPGKRVLVNGYRGNEGTLSEINEKAFSATIVITGPKLGRVRVEGIQYEDISKLA"

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 1386;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22

Db 1164 ATACCTTCAACTCTGCGTCCTT 1143

RESULT 5

AX003331/c 1390 bp DNA linear PAT 24-AUG-2000
 LOCUS
 DEFINITION Sequence 24 from Patent WO9929845.
 ACCESSION AX003331
 VERSION AX003331.1 GI:9927146

KEYWORDS

SOURCE Mus sp.

ORGANISM Mus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE

1 Angulo-Mora, J.F. and Mauffrey, P.

AUTHORS Sequences coding for kin17 protein and their applications

TITLE Patent: WO 9929845-A 24 17-JUN-1999;

JOURNAL ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE

(FR)

FEATURES

source Location/Qualifiers
 1..1390
 /organism="Mus sp."
 /mol_type="unassigned DNA"
 /db_xref="taxon:10095"

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 1390;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22

Db 1145 ATACCTTCAACTCTGCGTCCTT 1124

RESULT 6

MMKIN17/c 1414 bp mRNA linear ROD 30-OCT-1995
 LOCUS
 DEFINITION Mouse KIN17 mRNA for kin17 protein.
 ACCESSION X58472

VERSION X58472.1 GI:1045209

KEYWORDS KIN17 gene; kin17 protein; zinc-finger protein.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 1414)

Angulo, J.F., Rouer, E., Benarous, R. and Devoret, R.

AUTHORS Identification of a mouse cDNA fragment whose expressed polypeptide

TITLE reacts with anti-recA antibodies

JOURNAL Biochimie 73 (2-3), 251-256 (1991)

1715759

REFERENCE

2 Angulo, J.F., Rouer, E., Mazin, A., Mattei, M.G., Tissier, A.,

AUTHORS Horellou, P., Benarous, R. and Devoret, R.

TITLE Identification and expression of the cDNA of KIN17, a zinc-finger

gene located on mouse chromosome 2, encoding a new DNA-binding

JOURNAL protein

PUBMED Nucleic Acids Res. 19 (19), 5117-5123 (1991)

REFERENCE 1923796

AUTHORS Angulo, J.

TITLE Direct Submission

JOURNAL Submitted (25-MAR-1991) J. Angulo, C N R S, GEMC-Lab d'Enzymologie,

1 Avenue de la Terrasse, Gif-Sur-Yvette, Ile-de-France, FRANCE

REMARK revised by [4] MGPp

REFERENCE 4 (bases 1 to 1414)

AUTHORS Angulo, J.

TITLE Direct Submission

JOURNAL Submitted (30-OCT-1995) J. Angulo, C N R S, GEMC-Lab d'Enzymologie,

1 Avenue de la Terrasse, Gif-Sur-Yvette, Ile-de-France, FRANCE

COMMENT On Oct 31, 1995 this sequence version replaced gi:52795.

FEATURES

source Location/Qualifiers
 1..1414
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /cell_line="WB66"
 /tissue_type="brain"
 /dev_stage="embryo"

gene

1..1414

/gene="KIN17"

/pseudo

<1..1382

/gene="KIN17"

/pseudo

25..1200

/gene="KIN17"

/pseudo

/codon_start=1

/product="kin17 protein"

/db_xref="PSEUDO:CAA41386.1"

106..174

/gene="KIN17"

/note="zinc-finger of kin17"

/pseudo

234..842

/gene="KIN17"

/note="kin17(200) fragment"

/pseudo

/evidence=experimental

472..496

/gene="KIN17"

/note="kin17 helix1 motif"

/pseudo

781..792

/gene="KIN17"

/note="nuclear localization signal 1"

/pseudo

907..924

/gene="KIN17"

/note="nuclear localization signal 2"

/pseudo

1382

/gene="KIN17"

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 1414;

Best Local Similarity 100.0%; Pred. No. 2.8;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22

Db 1169 ATACCTTCAACTCTGCGTCCTT 1148


```

Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneses,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-MAR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2004 this sequence version replaced gi:24182248.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24315
Center clone name: 82_I_19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191018 bases at least Q40
Consensus quality: 191303 bases at least Q30
Consensus quality: 191388 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 191574; sum-of-contigs
Quality coverage: 12.2 in Q20 bases; agarose-fp
Quality coverage: 11.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 89963: contig of 89963 bp in length
* 89964 90063: gap of 100 bp
* 90064 94560: contig of 4497 bp in length
* 94561 94660: gap of 100 bp
* 94661 169601: contig of 74941 bp in length
* 169602 169701: gap of 100 bp
* 169702 180882: contig of 11181 bp in length.
-----
FEATURES
      source
      1..180882
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /clone="RP24-82119"
        /clone_lib="RPC1-24 Male Mouse BAC"
      1..89963
        /note="assembly_fragment
clone end:SP6
vector_side:left"
89964..90063
/estimated_length=100
90064..94560
/note="assembly_fragment"
94561..94660
/estimated_length=100
94661..169601
/estimated_length=100
-----
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneses,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-MAR-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2004 this sequence version replaced gi:24182248.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24315
Center clone name: 82_I_19
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 191018 bases at least Q40
Consensus quality: 191303 bases at least Q30
Consensus quality: 191388 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 191574; sum-of-contigs
Quality coverage: 12.2 in Q20 bases; agarose-fp
Quality coverage: 11.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 89963: contig of 89963 bp in length
* 89964 90063: gap of 100 bp
* 90064 94560: contig of 4497 bp in length
* 94561 94660: gap of 100 bp
* 94661 169601: contig of 74941 bp in length
* 169602 169701: gap of 100 bp
* 169702 180882: contig of 11181 bp in length.
-----
FEATURES
      source
      1..180882
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /clone="RP24-82119"
        /clone_lib="RPC1-24 Male Mouse BAC"
      1..89963
        /note="assembly_fragment
clone end:SP6
vector_side:left"
89964..90063
/estimated_length=100
90064..94560
/note="assembly_fragment"
94561..94660
/estimated_length=100
94661..169601
/estimated_length=100
-----

```

```

/note="assembly_fragment"
169602..169701
/estimated_length=100
169702..180882
/note="assembly_fragment
clone_end:T7
vector_side:right"
-----
Query Match      100.0%; Score 22; DB 14; Length 180882;
Best Local Similarity 100.0%; Pred. No. 2.3; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0;
QY 1 ATACCTTCAACTCTCGTCCTT 22
|||||
Db 46096 ATACCTTCAACTCTCGTCCTT 46117
|||||
RESULT 9
AC124010/c
LOCUS AC124010
DEFINITION Mus musculus chromosome UNK clone RP23-119N4, WORKING DRAFT
ACCESSION AC124010
VERSION AC124010.2 GI:22475963
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Aug 25, 2002 this sequence version replaced gi:21327639.
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BA0119N04
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 183528 bases at least Q40
Consensus quality: 184210 bases at least Q30
Consensus quality: 184756 bases at least Q20
Insert size: 197000; agarose-fp
Insert size: 196816; sum-of-contigs
Quality coverage: 9.53 in Q20 bases; agarose-fp
Quality coverage: 9.61 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 11868: contig of 11868 bp in length
 * 11869: gap of unknown length
 * 11969: contig of 16938 bp in length
 * 11969: gap of unknown length
 * 28907: gap of unknown length
 * 28907: contig of 43767 bp in length
 * 72774: gap of unknown length
 * 72874: contig of 52053 bp in length
 * 124927: gap of unknown length
 * 125026: contig of 56876 bp in length
 * 125027: gap of unknown length
 * 181903: contig of 398 bp in length.
 * 182003: contig of 398 bp in length.

FEATURES

source
 1. .182400
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="UNK"
 /clone="RP23-119N4"
 1. .11868
 /note="assembly_name:Contig86"
 11869. .11968
 /estimated_length=unknown
 11969. .28906
 /note="assembly_name:Contig87"
 28907. .29006
 /estimated_length=unknown
 29007. .72773
 /note="assembly_name:Contig88"
 72774. .72873
 /estimated_length=unknown
 72874. .124926
 /note="assembly_name:Contig89"
 124927. .125026
 /estimated_length=unknown
 125027. .181902
 /note="assembly_name:Contig90"
 181903. .182002
 /estimated_length=unknown
 182003. .182400
 /note="assembly_name:Contig16"

ORIGIN

Query Match 100.0%; Score 22; DB 14; Length 182400;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGTCCTT 22
 |||||
 Db 169317 ATACCTTCAACTCTCGTCCTT 169296

RESULT 10

AL772367
 LOCUS
 DEFINITION Mouse DNA sequence from clone RP23-119N4 on chromosome 2, complete sequence.
 ACCESSION AL772367
 VERSION AL772367.7 GI:22759512
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 185378)
 Direct Submission
 Submitted (30-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Sep 9, 2002 this sequence version replaced gi:22265488.

----- Genom Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC

Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-119N4 is from the RPCI-23 Mouse PAC Library constructed by the group of Pieter de Jong.
 For further details see http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6.

FEATURES

source

1. .185378
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP23-119N4"
 /clone_lib="RPCI-23"

ORIGIN

Query Match 100.0%; Score 22; DB 9; Length 185378;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 ATACCTTCAACTCTCGTCCTT 22
 |||||
 Db 139510 ATACCTTCAACTCTCGTCCTT 139531

RESULT 11

AC119716/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AC119716 221547 bp DNA linear HTG 19-NOV-2002
 Rattus norvegicus clone CH230-274116, *** SEQUENCING IN PROGRESS
 ***, 17 unordered pieces.

AC119716 GI:25092366

HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 221547)

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alibrooks, S., Amin, A., Anguiano, D.,

Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, I., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuewa, L., Louised, H., Lozada, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakoeleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiror, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, P., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajz, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, R., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 221547)
Worley, K.C.

REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (01-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 221547)

REFERENCE
AUTHORS
TITLE
JOURNAL

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23611286.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
----- Center project name: GWIC

Center clone name: CH230-274116
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 208013 bases at least Q40
Consensus quality: 211685 bases at least Q30
Consensus quality: 213907 bases at least Q20
Estimated insert size: 199182; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 198201: contig of 198201 bp in length
* 198202: gap of unknown length
* 198302: contig of 1448 bp in length
* 198302: gap of unknown length
* 199750: gap of unknown length
* 201115: contig of 1266 bp in length
* 201116: gap of unknown length
* 201215: contig of 1123 bp in length
* 202339: gap of unknown length
* 202339: contig of 1062 bp in length
* 203501: gap of unknown length
* 203501: contig of 1305 bp in length
* 204906: gap of unknown length
* 205006: contig of 1187 bp in length
* 206193: gap of unknown length
* 206293: contig of 1093 bp in length
* 207386: gap of unknown length
* 207485: contig of 1034 bp in length
* 208519: gap of unknown length
* 208520: contig of 1139 bp in length
* 208619: gap of unknown length
* 209759: gap of unknown length
* 209759: contig of 1583 bp in length
* 209859: gap of unknown length
* 211442: contig of 1583 bp in length
* 211542: gap of unknown length
* 212877: contig of 1335 bp in length
* 212977: gap of unknown length
* 214256: contig of 1280 bp in length
* 214357: gap of unknown length
* 214357: contig of 2647 bp in length
* 217003: gap of unknown length
* 217103: contig of 1356 bp in length
* 218459: gap of unknown length
* 218460: contig of 1229 bp in length
* 219788: gap of unknown length
* 219789: contig of 1659 bp in length.
* 219889: contig of 1659 bp in length.

FEATURES
Location/Qualifiers
1. .221547
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clones="CH230-274116"
1. .1272
/note="wgs end extension
clone end:Sp6"
complement(4416. .5263)
/note="clone boundary
clone end:Sp6
site:
end sequence:BZ266352"
192445. .193553
/note="wgs contig"
195743. .196592
/note="clone boundary
clone end:T7
site:
end_sequence:BZ266349"

misc_feature
misc_feature
misc_feature
misc_feature

```
gap      198202..198301
/estimated_length=unknown
gap      199750..199849
/estimated_length=unknown
gap      201116..201215
/estimated_length=unknown
gap      202339..202438
/estimated_length=unknown
gap      203501..203600
/estimated_length=unknown
gap      204906..205005
/estimated_length=unknown
gap      206193..206292
/estimated_length=unknown
gap      207386..207485
/estimated_length=unknown
gap      208520..208619
/estimated_length=unknown
gap      209759..209858
/estimated_length=unknown
gap      211442..211541
/estimated_length=unknown
gap      212877..212976
/estimated_length=unknown
gap      214257..214356
/estimated_length=unknown
gap      217004..217103
/estimated_length=unknown
gap      218460..218559
/estimated_length=unknown
gap      219789..219888
/estimated_length=unknown

Query Match      100.0%; Score 22; DB 14; Length 221547;
Best Local Similarity 100.0%; Pred. No. 2,3;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db      214156 ATACCTTCAACTCTGCGTCCTT 214135

RESULT 12
LOCUS      AX003328      22 bp      DNA      linear      PAT 24-AUG-2000
DEFINITION Sequence 21 from Patent WO9929845.
ACCESSION AX003328
VERSION    AX003328.1 GI:9927145
KEYWORDS   '
SOURCE     Homo sapiens (human)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE  1
AUTHORS   Angulo-Mora, J.F. and Mauffrey, P.
TITLE     Sequences coding for kin17 protein and their applications
JOURNAL   Patent: WO 9929845-A 21 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)
FEATURES   Location/Qualifiers
            source
            1..22
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
ORIGIN
Query Match      92.7%; Score 20.4; DB 6; Length 22;
Best Local Similarity 95.5%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db      914 ATCTCTTCAACTCTGCGTCCTT 893

RESULT 15
LOCUS      AX003311/c
DEFINITION Sequence 33 from Patent WO9929845.
ACCESSION AX003340
VERSION    AX003340.1 GI:9927153
KEYWORDS   '
SOURCE     Homo sapiens (human)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE  1
AUTHORS   Angulo-Mora, J.F. and Mauffrey, P.
TITLE     Sequences coding for kin17 protein and their applications
JOURNAL   Patent: WO 9929845-A 33 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)
FEATURES   Location/Qualifiers
            source
            1..333
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
ORIGIN
Query Match      92.7%; Score 20.4; DB 6; Length 333;
Best Local Similarity 95.5%; Pred. No. 21;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db      305 ATCTCTTCAACTCTGCGTCCTT 284

RESULT 14
LOCUS      AX003310/c
DEFINITION Sequence 3 from Patent WO9929845.
ACCESSION AX003310
VERSION    AX003310.1 GI:9927127
KEYWORDS   '
SOURCE     Homo sapiens (human)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE  1
AUTHORS   Angulo-Mora, J.F. and Mauffrey, P.
TITLE     Sequences coding for kin17 protein and their applications
JOURNAL   Patent: WO 9929845-A 3 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)
FEATURES   Location/Qualifiers
            source
            1..1002
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
ORIGIN
Query Match      92.7%; Score 20.4; DB 6; Length 1002;
Best Local Similarity 95.5%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db      914 ATCTCTTCAACTCTGCGTCCTT 893

RESULT 15
LOCUS      AX003311/c
DEFINITION Sequence 33 from Patent WO9929845.
ACCESSION AX003340
VERSION    AX003340.1 GI:9927153
KEYWORDS   '
SOURCE     Homo sapiens (human)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE  1
AUTHORS   Angulo-Mora, J.F. and Mauffrey, P.
TITLE     Sequences coding for kin17 protein and their applications
JOURNAL   Patent: WO 9929845-A 33 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)
FEATURES   Location/Qualifiers
            source
            1..333
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
ORIGIN
```

LOCUS AX003311 1002 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 4 from Patent WO929845.
ACCESSION AX003311
VERSION AX003311.1 GI:9927128
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Angulo-Mora, J. F. and Mauffrey, P.
TITLE Sequences coding for kin17 protein and their applications
JOURNAL Patent: WO 929845-A 4 17-JUN-1999;
ANGULO MORA JAIME FRANCISCO (FR); COMMISSARIAT ENERGIE ATOMIQUE
(FR)
FEATURES
source Location/Qualifiers
1..1002
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 92.7%; Score 20.4; DB 6; Length 1002;
Best Local Similarity 95.5%; Pred. No. 20;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATACCTTCAACTCTGCGTCCTT 22
||| ||||| ||||| ||||| |||||
Db 1002 ATCCTTCAACTCTGCGTCCTT 981

Search completed: November 26, 2005, 19:11:14
Job time : 119.932 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 12:48:57 ; Search time 14.1112 Seconds
(without alignments)
10390.557 Million cell updates/sec

Title: US-09-555-529-28

Perfect score: 22

Sequence: 1 ataccttcaactctgctgctctt 22

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	2	AAX85572 PCR prime
2	22	100.0	1102	2	AAX85550 CDNA of a
3	22	100.0	1390	2	AAX85570 CDNA of a
4	22	100.0	1458	2	AAX79936 Murine Ki
5	20.4	92.7	22	2	AAX85569 PCR prime
6	20.4	92.7	398	13	ACF87901 Human SIR
7	20.4	92.7	591	4	AAX84095 Human col
8	20.4	92.7	679	13	ADQ56943 Novel can
9	20.4	92.7	1002	2	AAX85552 Probe der
10	20.4	92.7	1002	2	AAX85551 CDNA of a
11	20.4	92.7	1296	2	AAX85549 CDNA of a
12	18.8	85.5	12332	10	AAD55664 Bovine vi
13	17.8	80.9	1432	13	ADX61400 Plant ful
14	17.8	80.9	1689	10	ACF68507 Photorhab
15	17.8	80.9	1921	13	ADX61280 Plant ful
16	17.8	80.9	110000	10	Continuation (13 o
17	17.8	80.9	182624	10	ACF67367 12
18	17.4	79.1	3882	11	ACL28061 Rice abio
19	17.4	79.1	11673	4	ABL20570 Drosophill

C	20	17.4	79.1	11707	4	ABL26444
	21	17.2	78.2	138	4	AAK79342
	22	17.2	78.2	339	10	ADI21316
C	23	17.2	78.2	825	8	ACA29919
	24	17.2	78.2	1820	10	ABZ24202
	25	17.2	78.2	1888	9	AAU62516
	26	17.2	78.2	2000	6	AAU36306
	27	17.2	78.2	2173	11	ADL33411
	28	17.2	78.2	2800	10	ABZ24200
	29	17.2	78.2	2818	10	ABZ24203
	30	17.2	78.2	2913	10	ADI21761
	31	17.2	78.2	3814	7	ADW31104
	32	17.2	78.2	3814	7	ADW41958
	33	17.2	78.2	3814	10	ABZ24201
	34	17.2	78.2	3814	12	ADO78126
	35	17.2	78.2	3876	6	AAU40003
C	36	17.2	78.2	111206	11	ACN45152
	37	17	77.3	1858	13	ADX32258
	38	17	77.3	1961	13	ADX55232
	39	17	77.3	1974	13	ADX55182
	40	16.8	76.4	498	9	ACC72960
	41	16.8	76.4	498	12	ADL09602
C	42	16.8	76.4	546	12	ADL11629
	43	16.8	76.4	986	13	ADX60098
C	44	16.8	76.4	1589	6	ABL35042
	45	16.8	76.4	2892	8	ADA69873

ALIGNMENTS

RESULT 1

AAX85572

ID AAX85572 standard; cDNA; 22 BP.

XX AAX85572;

DT 07-SEP-1999 (first entry)

DE PCR primer used to amplify human kin17 CDNA sequences.

XX Human; kin17 protein; cell proliferation; fertility;

KW hyperproliferative disease; protein interaction; curved DNA;

KW HIV replication; HIV integration; repair enzyme; PCR primer; ss.

XX Synthetic.

OS Mus sp.

XX FR2772046-A1.

XX PD 11-JUN-1999.

XX 09-DEC-1997; 97FR-00015536.

XX 09-DEC-1997; 97FR-00015536.

(COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX WPI; 1999-359999/31.

XX New DNA coding for human kin17 protein - useful for controlling cell

XX proliferation or fertility.

XX Example 1; Page 12; 69pp; French.

XX PCR primers AAX85571-72 were used to amplify 1000 base pairs of cDNA
XX encoding a human kin17 protein. The mammalian kin17 protein is useful for
XX preparing a medicament for controlling cell proliferation or for
XX controlling fertility. The medicaments can also be used to treat
XX hyperproliferative diseases. Fragments between amino acids 55 and 235
XX (preferably between amino acids 129 and 228) of a mammalian kin17 protein

CC are useful for regulating the interaction between proteins and curved
 CC DNA. The fragment can be used to block replication of HIV or its
 CC integration into the human genome or to target repair enzymes to curved
 CC DNA sites. Expression vectors for kin17 can be used for controlling cell
 CC proliferation

SQ Sequence 22 BP; 4 A; 8 C; 2 G; 8 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 2; Length 22;
 Best Local Similarity 100.0%; Pred. No. 1.5;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22
 |||||
 Db 1 ATACCTTCAACTCTGCGTCCTT 22

RESULT 2
 AAX85550/c
 ID AAX85550 standard; cDNA; 1102 BP.

XX

AC AAX85550;

XX

DT 07-SEP-1999 (first entry)

XX

DE cDNA of a gene coding for a mouse deleted kin17 protein.

XX kin17 protein; cell proliferation; fertility; hyperproliferative disease;
 KW protein interaction; curved DNA; HIV replication; HIV integration;
 KW repair enzyme; ss.
 XX

OS Mus sp.

PN FR2772046-A1.

XX

PD 11-JUN-1999.

XX

PF 09-DEC-1997; 97FR-00015536.

XX

PR 09-DEC-1997; 97FR-00015536.

XX

XX (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX

PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX

DR WPI; 1999-359999/31.

XX

PT New DNA coding for human kin17 protein - useful for controlling cell

XX

PT proliferation or fertility.

XX

PS Claim 4; Page 31; 69pp; French.

XX

XX The present sequence encodes a mouse kin17 protein with amino acids 129-

CC 228 deleted. The mammalian kin17 protein is useful for preparing a

CC medicament for controlling cell proliferation or for controlling

CC fertility. The medicaments can also be used to treat hyperproliferative

CC diseases. Fragments between amino acids 55 and 235 (preferably between

CC amino acids 129 and 228) of a mammalian kin17 protein are useful for

CC regulating the interaction between proteins and curved DNA. The fragment

CC can be used to block replication of HIV or its integration into the human

CC genome or to target repair enzymes to curved DNA sites. Expression

CC vectors for kin17 can be used for controlling cell proliferation

XX

SQ Sequence 1102 BP; 373 A; 205 C; 272 G; 252 T; 0 U; 0 Other;

Query Match 100.0%; Score 22; DB 2; Length 1102;
 Best Local Similarity 100.0%; Pred. No. 2.4;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22
 |||||
 Db 857 ATACCTTCAACTCTGCGTCCTT 836

RESULT 3

AAX85570/c

ID AAX85570 standard; cDNA; 1390 BP.

XX

AC AAX85570;

XX

DT 07-SEP-1999 (first entry)

XX

DE cDNA of a gene coding for the murine kin17 protein.

XX

KW Mouse; kin17 protein; cell proliferation; fertility;

KW hyperproliferative disease; protein interaction; curved DNA;

KW HIV replication; HIV integration; repair enzyme; ss.

XX

OS Mus sp.

XX

PN FR2772046-A1.

XX

PD 11-JUN-1999.

XX

PF 09-DEC-1997; 97FR-00015536.

XX

PR 09-DEC-1997; 97FR-00015536.

XX

PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

XX

PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;

XX

DR WPI; 1999-359999/31.

XX

PT New DNA coding for human kin17 protein - useful for controlling cell

XX

PT proliferation or fertility.

XX

PS Claim 21; Page 35-36; 69pp; French.

XX

XX The present sequence encodes a murine kin17 protein. The mammalian kin17

CC protein is useful for preparing a medicament for controlling cell

CC proliferation or for controlling fertility. The medicaments can also be

CC used to treat hyperproliferative diseases. Fragments between amino acids

CC 55 and 235 (preferably between amino acids 129 and 228) of a mammalian

CC kin17 protein are useful for regulating the interaction between proteins

CC and curved DNA. The fragment can be used to block replication of HIV or

CC its integration into the human genome or to target repair enzymes to

CC controlling cell proliferation

XX

SQ Sequence 1390 BP; 482 A; 256 C; 353 G; 299 T; 0 U; 0 Other;

XX

Query Match 100.0%; Score 22; DB 2; Length 1390;

Best Local Similarity 100.0%; Pred. No. 2.4;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22

Db 1145 ATACCTTCAACTCTGCGTCCTT 1124

XX

RESULT 4

AAQ79936/c

ID AAQ79936 standard; cDNA; 1458 BP.

XX

AC AAQ79936;

XX

DT 25-MAR-2003 (revised)

XX

DT 06-SEP-1995 (first entry)

XX

DE Murine Kin17 cDNA.

XX

KW chromosomal rearrangement; kin17 protein; SOS DNA repair system; RecA;

KW genotoxic agent; zinc finger; DNA binding protein; ss.

XX

OS Mus musculus.

```
XX FH Key Location/Qualifiers
FT primer_bind complement(1..21)
FT FT /*tag= b
FT FT /note= "Oligo L (AAQ79946) binds to complementary strand"
FT FT 22..1434
FT FT /*tag= 1
FT FT /label= kin17 cDNA
FT FT /note= "nucleotides 1-1414; the genomic DNA contains at
FT FT least two introns within this sequence, see Comments"
FT FT 32..149
FT FT /*tag= c
FT FT /note= "Oligo C (AAQ79938) binding site"
FT FT 46..1221
FT FT /*tag= a
FT FT /product= "Kin17"
FT FT /note= "N's in the sequence denote illegible residues"
FT FT complement(67..86)
FT FT primer_bind /*tag= d
FT FT /note= "Oligo S (AAQ79947) binds to complementary strand"
FT FT 274..1297
FT FT /*tag= e
FT FT /note= "Oligo D (AAQ79939) binding site"
FT FT complement(339..360)
FT FT primer_bind /*tag= f
FT FT /note= "Oligo K (AAQ79945) binds to complementary strand"
FT FT 451..1474
FT FT /*tag= g
FT FT /note= "Oligo J (AAQ79944) binding site"
FT FT complement(550..567)
FT FT primer_bind /*tag= h
FT FT /note= "Oligo E (AAQ79940) binds to complementary strand"
FT FT 802..1825
FT FT /*tag= i
FT FT /note= "Oligo F (AAQ79941) binding site"
FT FT complement(839..862)
FT FT primer_bind /*tag= j
FT FT /note= "Oligo G (AAQ79942) binds to complementary strand"
FT FT complement(1435..1458)
FT FT primer_bind /*tag= k
FT FT /note= "Oligo B (AAQ79937) binds to complementary strand"
FT FT
FT FR2706487-A1.
FT 23-DEC-1994.
FT 15-JUN-1993; 93FR-00007171.
FT 15-JUN-1993; 93FR-00007171.
FT (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
FT Angulo-Mora JF, Tissier A, Frelat G, Mauffrey P, Guilly M;
FT WPI; 1995-039031/06.
FT Purified murine kin17 protein prepn. for detecting chromosomal
FT rearrangements - also related antibodies, human and murine DNA, primers,
FT probes and vectors, used to assess damage caused by genotoxic agents.
FT Claim 9; Page 33; 54pp; French.
XX CC The murine Kin17 protein includes a zinc finger domain (see AAR66766),
XX CC recognises single- and double-stranded DNA (partic. regions of secondary
XX CC structure), has apparent mol. wt. 43 kD and is recognised by both anti-
XX CC kin17 antibodies and antibodies against the RecA protein of E.coli. The
XX CC kin17 protein is involved in DNA repair; it can be used to monitor
XX CC chromosomal rearrangements following exposure to genotoxic agents. The
XX CC kin17 cDNA sequence/AAQ79936 consists of a 1414 nucleotide sequence,
XX CC flanked by primer binding sites; the genomic kin17 DNA is claimed in
XX CC which an intron is inserted at position 137 of the 1414 nucleotide cDNA
XX CC and a second intron is located between nucleotides 339-429 of the cDNA.
XX CC (Updated on 25-MAR-2003 to correct PN field.)
```

```
XX SQ Sequence 1458 BP; 496 A; 272 C; 372 G; 311 T; 0 U; 7 Other;
XX Query Match 100.0%; Score 22; DB 2; Length 1458;
XX Best Local Similarity 100.0%; Pred. No. 2.4; Mismatches 0; Gaps 0;
XX Matches 22; Conservative 0; Indels 0; Gaps 0;
XX QY 1 ATACCTTCAACTCTGCGTCCTT 22
XX |||||
XX Db 1190 ATACCTTCAACTCTGCGTCCTT 1169
XX
XX RESULT 5
XX AAX85569
XX ID AAX85569 standard; cDNA; 22 BP.
XX AC AAX85569;
XX DT 07-SEP-1999 (first entry)
XX DE PCR primer and probe used to detect human kin17 nucleic acids.
XX KW Human; kin17 protein; cell proliferation; fertility; probe;
XX KW hyperproliferative disease; protein interaction; curved DNA;
XX KW HIV replication; HIV integration; repair enzyme; PCR primer; ss.
XX OS Synthetic.
XX OS Homo sapiens.
XX PN FR2772046-A1.
XX PD 11-JUN-1999.
XX PF 09-DEC-1997; 97FR-00015536.
XX PR 09-DEC-1997; 97FR-00015536.
XX PA (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.
XX PI Kannouche P, Mauffrey P, Pinon LG, Biard D, Angulo MJF;
XX WPI; 1999-359999/31.
XX PT New DNA coding for human kin17 protein - useful for controlling cell
XX PT proliferation or fertility.
XX PS Claim 7; Page 5; 69pp; French.
XX CC AAX85553-69 represent PCR primers and probes used to detect nucleic acids
XX CC encoding human kin17 protein. The mammalian kin17 protein is useful for
XX CC preparing a medicament for controlling cell proliferation or for
XX CC controlling fertility. The medicaments can also be used to treat
XX CC hyperproliferative diseases. Fragments between amino acids 55 and 235
XX CC (preferably between amino acids 129 and 228) of a mammalian kin17 protein
XX CC are useful for regulating the interaction between proteins and curved
XX CC DNA. The fragment can be used to block replication of HIV or its
XX CC integration into the human genome or to target repair enzymes to curved
XX CC DNA sites. Expression vectors for kin17 can be used for controlling cell
XX CC proliferation
XX SQ Sequence 22 BP; 3 A; 8 C; 2 G; 9 T; 0 U; 0 Other;
XX Query Match 92.7%; Score 20.4; DB 2; Length 22;
XX Best Local Similarity 95.5%; Pred. No. 8.7; Indels 0; Gaps 0;
XX Matches 21; Conservative 0; Mismatches 1;
XX QY 1 ATACCTTCAACTCTGCGTCCTT 22
XX |||||
XX Db 1 ATTCCTTCAACTCTGCGTCCTT 22
XX
XX RESULT 6
XX ACF87901
```


Db	7433	ATTCCTTCAACTCTGAGTCCTT	7412	Best Local Similarity	90.5%;	Pred. No. 2.3e+02;	Mismatches	2;	Indels	0;	Gaps	0;
				Matches	19;	Conservative	0;					
RESULT 13												
ADX61400/c												
ID	ADX61400	standard; cDNA; 1432 BP.										
XX	ADX61400;											
AC												
XX												
XX												
DT	21-APR-2005	(first entry)										
XX												
DE	Plant full length insert polynucleotide seqid 32243.											
XX												
KW	plant protectant; plant growth regulant; gene therapy; plant;											
KW	recombinant DNA construct; physical array; plant breeding marker;											
KW	cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;											
KW	extreme osmotic condition; pathogen tolerance; pest tolerance;											
KW	growth rate; cell cycle pathway; disease resistance;											
KW	galactomannan production; lignin production; plant growth regulator;											
KW	yield; plant growth; plant development; seed oil; protein yield;											
KW	protein content; gene; ss.											
XX												
OS	Unidentified.											
XX												
XX	US2004034888-A1.											
PN												
XX												
PD	19-FEB-2004.											
XX												
XX	28-APR-2003; 2003US-00425114.											
XX												
PF	06-MAY-1999; 99US-00304517.											
PR	03-NOV-2001; 2001US-00985678.											
XX												
XX	(LIU//) LIU J.											
PA	(ZHOU//) ZHOU Y.											
PA	(KOVA//) KOVALIC D K.											
PA	(SCRE//) SCREEN S E.											
PA	(TAB//) TABASKA J E.											
PA	(CAOY//) CAO Y.											
XX												
XX	Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;											
PI	WPI; 2004-180133/17.											
XX												
DR												
XX												
PT	New recombinant DNA construct, useful for improving plant tolerance to											
PT	cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or											
PT	pests, for conferring increased resistance to plant disease, or for											
PT	improving yield.											
XX												
PS	Claim 1; SEQ ID NO 32243; 15pp; English.											
XX												
CC	The invention describes a recombinant DNA construct comprising a											
CC	polynucleotide consisting of a sequence encoding an amino acid sequence											
CC	available in electronic form from the US patent office at											
CC	ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide											
CC	of the invention are also useful in physical arrays of molecules and as											
CC	plant breeding markers. The recombinant DNA construct is useful for											
CC	improving plant tolerance to cold, heat, drought, herbicides, extreme											
CC	osmotic conditions, pathogens or pests, for manipulating growth rate in											
CC	plant cells by modification of the cell cycle pathway, for conferring											
CC	increased resistance to plant disease, for producing galactomannan,											
CC	lignin or plant growth regulators, for increasing the rate of homologous											
CC	recombination in plants, for improving yield by modification of											
CC	photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake											
CC	or by providing improved plant growth and development under at least one											
CC	stress condition or for modifying seed oil or protein yield and/or											
CC	content. This sequence represents a plant full length insert											
CC	polynucleotide that can be used in the recombinant DNA construct of the											
CC	invention.											
XX												
XX	Sequence 1432 BP; 406 A; 294 C; 316 G; 416 T; 0 U; 0 Other;											
XX												
Query Match	80.9%;	Score 17.8;	DB 13;	Length 1432;								
Best Local Similarity	90.5%;	Pred. No. 2.4e+02;										
Matches	19;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;			

QY 2 TACCTTCAACTCTGGTCCTT 22
DB 1333 TACCTTCAACTCTTTGTCTT 1313

RESULT 15
AD61280/c
ID AD61280 standard; cDNA; 1921 BP.
XX AC AD61280;
XX DT 21-APR-2005 (first entry)
XX DE Plant full length insert polynucleotide seqid 32123.
XX KW plant protectant; plant growth regulant; gene therapy; plant;
KW recombinant DNA construct; physical array; plant breeding marker;
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;
KW extreme osmotic condition; pathogen tolerance; pest tolerance;
KW growth rate; cell cycle pathway; disease resistance;
KW galactomannan production; lignin production; plant growth regulator;
KW yield; plant growth; plant development; seed oil; protein yield;
KW protein content; gene; ss.
XX OS Unidentified.
XX PN US2004034888-A1.
XX PD 19-FEB-2004.
XX PF 28-APR-2003; 2003US-00425114.
XX PR 06-MAY-1999; 99US-00304517.
XX PR 05-NOV-2001; 2001US-00985678.
XX PA (LIUJ/) LIU J.
XX PA (ZHOU/) ZHOU Y.
XX PA (KOVA/) KOVALIC D K.
XX PA (SCRE/) SCREEN S E.
XX PA (TABAS/) TABASKA J E.
XX PA (CAOY/) CAO Y.
XX PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX WPI; 2004-180133/17.
XX PT New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX PS Claim 1; SEQ ID NO 32123; 15pp; English.
XX CC The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX

SQ Sequence 1921 BP; 491 A; 455 C; 466 G; 509 T; 0 U; 0 Other;
Query Match 80.9%; Score 17.8; DB 13; Length 1921;
Best Local Similarity 90.5%; Pred. NO. 2.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TACCTTCAACTCTGGTCCTT 22
DB 1628 TTCCTTCAGCTCTGGTCCTT 1608

Search completed: November 26, 2005, 13:40:56
Job time : 16.1112 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 26, 2005, 12:58:53 ; Search time 122.977 Seconds
(without alignments)
8370.011 Million cell updates/sec

Title: US-09-555-529-28
Perfect score: 22
Sequence: 1 ataccttcgaactctgcgtcctt 22

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

- 1: gb_est1.*
- 2: gb_est2.*
- 3: gb_est3.*
- 4: gb_hic.*
- 5: gb_est4.*
- 6: gb_est5.*
- 7: gb_est6.*
- 8: gb_est7.*
- 9: gb_gsa1.*
- 10: gb_gsa2.*
- 11: gb_gsa3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	22	100.0	238	1	AV244074
C 2	22	100.0	272	1	AV091267
C 3	22	100.0	283	7	CK334767
4	22	100.0	322	1	A1462821
5	22	100.0	391	1	AU041571
C 6	22	100.0	399	5	BY395138
C 7	22	100.0	430	6	CF617307
C 8	22	100.0	447	2	BB749127
9	22	100.0	457	1	AW045702
C 10	22	100.0	466	2	B2448684
11	22	100.0	512	3	BM228037
12	22	100.0	512	3	BM229452
13	22	100.0	513	2	BE225867
14	22	100.0	560	2	BG072126
15	22	100.0	601	3	BP774644
16	22	100.0	677	2	BE321724
C 17	22	100.0	1081	2	BE283156
C 18	22	100.0	1154	10	AV412520
C 19	20.4	92.7	202	3	BQ307881
C 20	20.4	92.7	238	2	BF456707
C 21	20.4	92.7	242	6	CB243501
C 22	20.4	92.7	244	1	AV239958

C 23	20.4	92.7	259	6	CF526711
C 24	20.4	92.7	328	2	BE091046
C 25	20.4	92.7	348	1	AI261989
26	20.4	92.7	384	1	AA770446
27	20.4	92.7	386	3	BM661987
28	20.4	92.7	398	8	H75516
C 29	20.4	92.7	437	2	BB749400
C 30	20.4	92.7	469	1	AI087818
31	20.4	92.7	476	1	AI378396
32	20.4	92.7	485	1	AI089251
C 33	20.4	92.7	488	3	BM689918
C 34	20.4	92.7	512	5	BG548883
35	20.4	92.7	516	2	BQ944189
36	20.4	92.7	521	6	BE349628
C 37	20.4	92.7	543	6	CB158644
38	20.4	92.7	558	5	BU737321
C 39	20.4	92.7	575	3	BM750112
C 40	20.4	92.7	591	1	AJ397270
C 41	20.4	92.7	592	2	BE896845
42	20.4	92.7	592	8	CX866354
43	20.4	92.7	604	9	AQ342366
44	20.4	92.7	605	6	CB242699
C 45	20.4	92.7	660	1	AL597250

ALIGNMENTS

RESULT 1
AV244074/c
LOCUS
DEFINITION
musculus cDNA clone 4831427J19 3' similar to X58472 Mouse KIN17
mRNA for kin17 protein, mRNA sequence.

ACCESSION
AV244074.1 GI:6231533
VERSION
EST.
KEYWORDS
SOURCE
Mus musculus (house mouse)

REFERENCE
AUTHORS

1 (bases 1 to 238)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateno, M., Tomimatsu, N., Tsunoda, Y., Watanabe, S., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. RIKEN Mouse ESTs (Konno, H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-resgsc.riken.jp, URL: http://genome.gsc.riken.jp/
Sasaki, N., Izawa, M., Watanabe, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

TITLE
JOURNAL
COMMENT

Transcriptional sequencing: A method for DNA sequencing using RNA polymers. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)


```

libraries"
/lab_host="DH10B"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/notes="Vector: pSPOT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearranged set of 15,247 clones from 11
embryo cDNA libraries (including preimplantation stage
embryos from unfertilized egg to blastocyst, embryonic
part of E7.5 embryos, extraembryonic part of E7.5
embryos, and E12.5 female mesonephros/sonad) and one
newborn ovary cDNA library. Average insert size 1.5 kb.
All source libraries are cloned unidirectionally with
Oligo(dT)-Not primers. References include: (1)
Genome-wide expression profiling of mid-gestation
placenta and embryo using a 15,000 mouse developmental
cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97:
9127-9132; (2) Large-scale cDNA analysis reveals phased
gene expression patterns during preimplantation mouse
development, 2000, Development, 127: 1737-1749; (3)
Genome-wide mapping of unselected transcripts from
extraembryonic tissue of 7.5-day mouse embryos reveals
enrichment in the t-complex and under-representation on
the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN
Query Match 100.0%; Score 22; DB 7; Length 283;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db 215 ATACCTTCAACTCTGCGTCCTT 236

RESULT 4
AI462821
LOCUS AI462821 322 bp mRNA linear EST 09-MAR-1999
DEFINITION v462821.x1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
IMAGE:752229 3', mRNA sequence.
ACCESSION AI462821
VERSION AI462821.1 GI:4316851
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 322)
Marra M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Waller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
Contact: Maria M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MG1:461213
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
High quality sequence stop: 315.
FEATURES
Location/Qualifiers
1..322
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:752229"

libraries"
/lab_host="DH10B"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/notes="Vector: pSPOT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearranged set of 15,247 clones from 11
embryo cDNA libraries (including preimplantation stage
embryos from unfertilized egg to blastocyst, embryonic
part of E7.5 embryos, extraembryonic part of E7.5
embryos, and E12.5 female mesonephros/sonad) and one
newborn ovary cDNA library. Average insert size 1.5 kb.
All source libraries are cloned unidirectionally with
Oligo(dT)-Not primers. References include: (1)
Genome-wide expression profiling of mid-gestation
placenta and embryo using a 15,000 mouse developmental
cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97:
9127-9132; (2) Large-scale cDNA analysis reveals phased
gene expression patterns during preimplantation mouse
development, 2000, Development, 127: 1737-1749; (3)
Genome-wide mapping of unselected transcripts from
extraembryonic tissue of 7.5-day mouse embryos reveals
enrichment in the t-complex and under-representation on
the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN
Query Match 100.0%; Score 22; DB 1; Length 322;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db 210 ATACCTTCAACTCTGCGTCCTT 231

RESULT 5
AU041571
LOCUS AU041571 391 bp mRNA linear EST 04-DEC-1998
DEFINITION AU041571 Mouse four-cell-embryo cDNA Mus musculus cDNA clone
J1006B10 3', mRNA sequence.
ACCESSION AU041571
VERSION AU041571.1 GI:3955806
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 391)
Ko, M.S.H., Kitchen, J.R., Wang, X., Threat, T.A., Sun, T.,
Depalma, G.E., Liang, Y., Kargul, G.J., Sharara, R., Lim, M.K. and
Doi, H.
Systematic analyses of genes expressed in 4-cell mouse embryo (The
ERATO/Doi Project at Wayne State University)
Unpublished (1998)
Contact: Hirofumi Doi
Doi Biosymmetry Project, ERATO
Japan Science and Technology Corporation (JST)
WBG Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdo@bioa.jst.go.jp.
FEATURES
Location/Qualifiers
1..391
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="J1006B10"
/dev_stage="four-cell-embryo"
/clone_lib="Mouse four-cell-embryo cDNA"

ORIGIN
Query Match 100.0%; Score 22; DB 1; Length 391;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
|||||
Db 215 ATACCTTCAACTCTGCGTCCTT 236

RESULT 6
BY395138/c

```

LOCUS	BY395138	399 bp	mRNA	linear	EST 12-DEC-2002
DEFINITION	BY395138 RIKEN full-length enriched, CRL-1722 L5178Y-R Mus musculus CDNA clone I730004D10 3', mRNA sequence.				
ACCESSION	BY395138				
VERSION	BY395138.1				
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidae; Muridae; Murinae; Mus.				
	1 (bases 1 to 399)				
AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaide, I., Oeato, N., Saito, R., Saito, K., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierzki, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.				
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
	Nature 420, 563-573 (2002)				
JOURNAL	12466851				
PUBLISHED	Contact: Yoshihide Hayashizaki				
COMMENT	Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute				
	The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/ Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a non-redundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.				
FEATURES	source				
	1..399				
FEATURES	Location/Qualifiers				
	/organism="Mus musculus"				
FEATURES	/mol_type="mRNA"				
	/strain="DBA/2"				
FEATURES	/db_xref="taxon:10090"				
	/clone="I730004D10"				
FEATURES	/cell_line="CRL-1722 L5178Y-R"				
	/clone_lib="RIKEN full-length enriched, CRL-1722 L5178Y-R"				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap-frozen and kept at -80C for two days before RNA extraction and purification (Tri-reagent method). cDNA was primed using oligo-dT primer."				
	5'-pGACTACTCTAGTCGCGAGCGGCC(T)25-3' and cloned into the EcorV/NotI sites of pExpress-1. Size-selection >0.75kb resulted in an average insert size of 1.1 kb. This primary, nanoquantity library is normalized to Cot5 (non-normalized primary library is NIH MGC 223) and was constructed by Express Genomics (Frederick, MD)."				
FEATURES	Location/Qualifiers				
	1..430				
FEATURES	/organism="Mus musculus"				
	/mol_type="mRNA"				
FEATURES	/db_xref="taxon:10090"				
	/clone="IMAGE:30526444"				
FEATURES	/lab_host="DH10B (phage-resistant)"				
	/clone_lib="NIH MGC_204"				
FEATURES	/note="Organ: placenta; Vector: pExpress-1; Site 1: EcorV; Site 2: NotI; RNA obtained from three placentas from female C57/BL6 mouse at 16 days pregnancy. Tissues were snap				

ORIGIN

Query Match 100.0%; Score 22; DB 6; Length 430;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGTCTT 22
 |||||
 DB 185 ATACCTTCAACTCTGGTCTT 164

RESULT 8

BB749127/c
 LOCUS BB749127 RIKEN full-length enriched, pooled tissues, pituitary,
 DEFINITION etc. Mus musculus cDNA clone G030102N09 3', mRNA sequence.

ACCESSION BB749127

VERSION BB749127.1 GI:16153363

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidae; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 447)
 Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
 Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
 Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
 Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
 Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
 Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
 Tanaka,T., Tomaru,A., Toyota,T., Watanishi,A., Yasunishi,A.,
 Muramatsu,M. and Hayashizaki,Y.

RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
 2001)

JOURNAL

Unpublished (2001)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Teurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-res@gsr.riken.jp, URL: http://genome.gsc.riken.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 wagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
 Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
 and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
 Sugahara,Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (http://genome.gsc.riken.go.jp) for
 further details.

e mouse tissues.

FEATURES

source

1. Location/Qualifiers

1. 447
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G030102N09"
 /clone_lib="RIKEN full-length enriched, pooled tissues,

pituitary, etc."

/note="pooled tissues ; (tissue_type=pituitary gland,
 dev_stage=adult, sex=male), (tissue_type=thymus,
 dev_stage=adult, sex=male), (tissue_type=forelimb,
 dev_stage=13 days embryo, sex=mixed), (tissue_type=testis,
 dev_stage=13 days embryo, sex=male),
 (tissue_type=olfactory brain, dev_stage=adult, sex=male)"

ORIGIN

Query Match 100.0%; Score 22; DB 2; Length 447;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGTCTT 22
 |||||
 DB 234 ATACCTTCAACTCTGGTCTT 213

RESULT 9

AW045702

LOCUS AW045702

DEFINITION UI-M-BHI-akq-d-04-0-UI-el NIH BMAP M_S2 Mus musculus cDNA clone
 AW045702 457 bp mRNA linear EST 18-SEP-1999

ACCESSION AW045702

VERSION AW045702.1 GI:5906231

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 Sciurognathi; Muroidae; Muridae; Murinae; Mus.

1 (bases 1 to 457)

Bonaldo,M.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

8889548

JOURNAL

PUBMED

COMMENT

Contact: Chin, H

National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD

20892-9643, USA

Tel: 301 443 1706

Fax: 301 443 9890

Email: m85t@mail.nih.gov

The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 normalized corpus striatum library cDNA library preparation: M.B.
 Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
 available by the means that is soon to be determined. When NIH
 determines the means for distribution of the BMAP cDNA clones, this
 record will be updated accordingly when that means is determined.
 The following repetitive elements were found in this cDNA sequence:
 1-21. >AT richlow complexity
 Seq primer: M13 Forward

POLYA=Yes.

FEATURES

source

1. Location/Qualifiers

1. 457
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BHI-akq-d-04-0-UI"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NIH BMAP M_S2"
 /note="vector: pT73D-Pac (pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; The
 NIH BMAP M_S2 library is a subtracted library derived from
 NIH BMAP M_S1, which in turn is a subtracted library
 derived from a mixture of normalized libraries from ten


```

ORIGIN
Query Match      100.0%; Score 22; DB 3; Length 512;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22
    |||||
Db 134 ATACCTTCAACTCTGCGTCCTT 155

RESULT 12
BM229452
LOCUS
DEFINITION
BM229452.2 GI:31486939
VERSION
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
Construction of long-transcript enriched cDNA libraries from
submicrogram amounts of total RNAs by a universal PCR amplification
method
Genome Res. 11 (9), 1553-1558 (2001)
11544199
On Dec 14, 2001 this sequence version replaced gi:17792386.
Other ESTs: K0278H05-5N
Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdaelgaun.grc.nia.nih.gov
Plate: K0278 row: H column: 05
Seq primer: -21M13 Forward
High quality sequence stop: 512
POLYA=Yes.

FEATURES
Location/Qualifiers
1..512
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:K0278H05-3"
/db_xref="taxon:10090"
/clone="NIA:K0278H05 IMAGE:30052216"
/tissue_type="Unfertilized Egg"
/lab_host="DH10B"
/clone_lib="NIA Mouse Unfertilized Egg cDNA Library
(Long)"
/notes="vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
extracted from a pool of 1488 unfertilized eggs.
Double-stranded cDNAs were synthesized with an Oligo(dT)
primer [Invitrogen]:
5'-pGAGTAGTCTTAGATCGAGCGCGCCCTTTTTTTTTTTT-3'),
treated with T4 DNA polymerase, and purified by
ethanol-precipitation. The cDNAs were ligated to
long-linker L1-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.

ORIGIN
Query Match      100.0%; Score 22; DB 3; Length 512;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCTT 22
    |||||
Db 134 ATACCTTCAACTCTGCGTCCTT 155

RESULT 13
BE225867
LOCUS
DEFINITION
BE225867.1 GI:8931103
VERSION
KEYWORDS
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
Wyllie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B.,
Ritter,B., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M.,
McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and
Bowers,Y.
WashU-Harvard Pancreas EST Project
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Adimika Meadows
(meadows@fas.harvard.edu)
MGI:1847913 This sequence now available from the IMAGE consortium,
for clone orders contact: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 384.

FEATURES
Location/Qualifiers
1..513
/organism="Mus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:563889"
/sex="Both"
/tissue_type="Pancreatic Bud"
/dev_stage="Embryonic day 10.5 and 12.5, mixed"
/lab_host="DH10B"
/clone_lib="Mouse E10 5 12 5 Pancreas cDNA Library"
/notes="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal 1; Library constructed using Superscript
Plasmid Library kit (Life Technologies). cDNA made by
oligo-dT priming. Size-selected by column fractionation;
average insert size 1.47 Kb. Primary library, unamplified.
cDNA Library Preparation: Guolin Chen."

ORIGIN

```

The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.5 kb. The library was constructed by Yulan Piao (NIA)."

```

Query Match      100.0%; Score 22; DB 2; Length 513;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGTCCTT 22
|||||
Db 208 ATACCTTCAACTCTCGTCCTT 229
|||||

RESULT 14
BG072126          560 bp mRNA linear EST 18-DEC-2003
LOCUS
DEFINITION
H3107A09-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
BG072126
BG072126.2 GI:40072826
EST.
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 560)
Tanaka,T.S., Jaradat,S.A., Lim,M.K., Kargul,G.J., Wang,X.,
Grahovac,M.J., Pantano,S., Sano,Y., Piao,Y., Negaraja,R., Doi,H.,
Wood,W.H. III, Becker,K.G. and Ko,M.S.H.
Genome-wide expression profiling of mid-gestation placenta and
embryo using a 15,000 mouse developmental cDNA microarray
Proc. Natl. Acad. Sci. U.S.A. 97 (16), 9127-9132 (2000)
10922068
On Jan 26, 2001 this sequence version replaced gi:12554695.
Other ESTs: H3107A09-5
Contact: George J. Kargul
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Casell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
Plate: H3107 row: A column: 09
Seq primer: -21M13 Forward
High quality sequence stop: 560
POLYA=Yes.

FEATURES
source
1..560
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="niaEST:H3107A09-3"
/clone="H3107A09"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/notes="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals
enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

ORIGIN
Query Match      100.0%; Score 22; DB 2; Length 560;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGTCCTT 22
|||||
Db 215 ATACCTTCAACTCTCGTCCTT 236
|||||

RESULT 15
BP774644          601 bp mRNA linear EST 02-DEC-2004
LOCUS
DEFINITION
BP774644 mouse (C57BL/6) pancreatic islet library with
recombination-based method Mus musculus cDNA clone mif14029 3',
mRNA sequence.
ACCESSION
BP774644
VERSION
BP774644.1 GI:50233342
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 601)
Nishimura,M., Yokoi,N., Miki,T., Horikawa,Y., Yoshioka,H.,
Takeda,J., Ohara,O. and Seino,S.
Construction of a multi-functional cDNA library specific for mouse
pancreatic islets and its application to microarray
DNA Res. 11 (5), 315-323 (2004)
15747579
Contact: Susumu Seino
Division of Cellular and Molecular Medicine
Kobe University Graduate School of Medicine
7-5-1 Kusunoki-cho, Chuo-ku, Kobe, Hyogo 650-0017, Japan
Tel: 81-78-382-5360
Fax: 81-78-382-5370
Email: seino@med.kobe-u.ac.jp.

FEATURES
source
1..601
Location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="mif14029"
/sex="male"
/tissue_type="pancreatic islet"
/dev_stage="adult"
/clone_lib="mouse (C57BL/6) pancreatic islet library with
recombination-based method"

ORIGIN
Query Match      100.0%; Score 22; DB 3; Length 601;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGTCCTT 22
|||||
Db 244 ATACCTTCAACTCTCGTCCTT 265
|||||

Search completed: November 27, 2005, 00:58:00
Job time : 124.977 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:02 ; Search time 4.556 Seconds
(without alignments)
8583.479 Million cell updates/sec

Title: US-09-555-529-28

Perfect score: 22

Sequence: 1 ataccttcaactctgcgtcctt 22

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2 6/ptodata/1/ina/1 COMB.seq:*
- 2: /cgn2 6/ptodata/1/ina/5 COMB.seq:*
- 3: /cgn2 6/ptodata/1/ina/6A COMB.seq:*
- 4: /cgn2 6/ptodata/1/ina/6B COMB.seq:*
- 5: /cgn2 6/ptodata/1/ina/H COMB.seq:*
- 6: /cgn2 6/ptodata/1/ina/PCTUS COMB.seq:*
- 7: /cgn2 6/ptodata/1/ina/PP COMB.seq:*
- 8: /cgn2 6/ptodata/1/ina/RE COMB.seq:*
- 9: /cgn2 6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.8	85.5	675	3	US-09-248-796A-2146
C 2	17.2	78.2	601	3	US-09-949-016-145512
C 3	17.2	78.2	601	3	US-09-949-016-145513
C 4	17.2	78.2	2798	3	US-09-949-016-4106
C 5	17.2	78.2	26103	3	US-09-949-016-16841
C 6	17.2	78.2	89450	3	US-09-949-016-15848
C 7	16.8	76.4	15543	3	US-09-949-016-17225
C 8	16.4	74.5	1988	2	US-08-684-862-10
C 9	16.4	74.5	1333	2	US-08-684-862-9
C 10	16.2	73.6	601	3	US-09-949-016-71590
C 11	16.2	73.6	601	3	US-09-949-016-179033
C 12	16.2	73.6	601	3	US-09-949-016-179034
C 13	16.2	73.6	703	3	US-09-270-767-1498
C 14	16.2	73.6	703	3	US-09-270-767-16780
C 15	16.2	73.6	3698	3	US-09-976-594-1098
C 16	16.2	73.6	6133	3	US-09-453-702B-15
C 17	16.2	73.6	6133	3	US-10-114-170-15
C 18	16.2	73.6	11835	3	US-09-949-016-16904
C 19	16.2	73.6	46885	3	US-09-949-016-13848
C 20	16.2	73.6	108310	3	US-09-949-016-16366
C 21	16.2	73.6	536165	3	US-09-214-808-1
C 22	16	72.7	27600	3	US-09-949-016-15290
C 23	15.8	71.8	601	3	US-09-949-016-183894
C 24	15.8	71.8	732	3	US-09-248-796A-6535

C 25	15.8	71.8	2194	3	US-09-949-016-2036	Sequence 2036, Ap
C 26	15.8	71.8	18857	3	US-09-949-016-13778	Sequence 13778, A
C 27	15.8	71.8	19038	3	US-09-659-737A-11	Sequence 11, Appl
C 28	15.8	71.8	19038	3	US-10-885-921-11	Sequence 11, Appl
C 29	15.8	71.8	450395	3	US-09-949-016-15473	Sequence 15473, A
C 30	15.6	70.9	316	3	US-09-533-559-3952	Sequence 3952, Ap
C 31	15.6	70.9	325	3	US-08-956-171E-4329	Sequence 4329, Ap
C 32	15.6	70.9	325	3	US-08-781-986A-4329	Sequence 4329, Ap
C 33	15.6	70.9	396	3	US-09-513-999C-24163	Sequence 24163, A
C 34	15.6	70.9	792	3	US-09-328-352-1107	Sequence 1107, Ap
C 35	15.6	70.9	1026	3	US-09-248-796A-2492	Sequence 2492, Ap
C 36	15.6	70.9	1074	3	US-09-107-433-419	Sequence 419, App
C 37	15.6	70.9	1128	3	US-09-583-110-2467	Sequence 2467, Ap
C 38	15.6	70.9	1365	3	US-09-578-063-34	Sequence 34, Appl
C 39	15.6	70.9	1980	3	US-09-578-063-33	Sequence 33, Appl
C 40	15.6	70.9	2139	3	US-09-614-221A-417	Sequence 417, App
C 41	15.6	70.9	3648	3	US-09-614-221A-233	Sequence 233, App
C 42	15.6	70.9	3675	3	US-10-104-047-277	Sequence 277, App
C 43	15.6	70.9	4591	3	US-09-976-594-1082	Sequence 1082, Ap
C 44	15.6	70.9	7158	3	US-09-543-681A-2132	Sequence 2132, Ap
C 45	15.6	70.9	11890	3	US-09-949-016-13794	Sequence 13794, A

ALIGNMENTS

RESULT 1

US-09-248-796A-2146/c
; Sequence 2146, Application US/09248796A
; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248, 796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074, 725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096, 409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 2146

; LENGTH: 675

; TYPE: DNA

; ORGANISM: Candida albicans

; US-09-248-796A-2146

Query Match 85.5%; Score 18.8; DB 3; Length 675;

Best Local Similarity 90.9%; Pred. No. 17;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22

Db 116 AGACCTTCAACTCTGCGTCCTT 95

RESULT 2

US-09-949-016-145512/c

; Sequence 145512, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949, 016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241, 755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237, 768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231, 498

```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 145512
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-145512

Query Match      78.2%; Score 17.2; DB 3; Length 601;
Best Local Similarity 86.4%; Pred. No. 96;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGTGCCTT 22
   ||||| ||||| ||||| |||||
Db 130 ATACCTTCTACTCTGTGTACTT 109

RESULT 3
US-09-949-016-145513/c
; Sequence 145513, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 145513
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-145513

Query Match      78.2%; Score 17.2; DB 3; Length 601;
Best Local Similarity 86.4%; Pred. No. 96;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGTGCCTT 22
   ||||| ||||| ||||| |||||
Db 415 ATACCTTCTACTCTGTGTACTT 394

RESULT 4
US-09-949-016-4106
; Sequence 4106, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4106
; LENGTH: 2798
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4106

Query Match      78.2%; Score 17.2; DB 3; Length 2798;
Best Local Similarity 86.4%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGTGCCTT 22
   ||||| ||||| ||||| |||||
Db 1572 ATACCTTCTACTCTGTGTACTT 1593

RESULT 5
US-09-949-016-16841/c
; Sequence 16841, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16841
; LENGTH: 26103
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(26103)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16841

Query Match      78.2%; Score 17.2; DB 3; Length 26103;
Best Local Similarity 86.4%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGGTGCCTT 22
   ||||| ||||| ||||| |||||
Db 16895 AAAGCTTCTACTCTGGTGCCTT 16874

RESULT 6
US-09-949-016-15848
; Sequence 15848, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15848
; LENGTH: 89450
; TYPE: DNA
```

```
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(89450)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15848

Query Match          78.2%; Score 17.2; DB 3; Length 89450;
Best Local Similarity 86.4%; Pred. No. 1.9e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGTCCTT 22
Db 84421 ATACCTTCTACTCTGTGTACTT 84442

RESULT 7
US-09-949-016-17225
; Sequence 17225, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17225
; LENGTH: 15543
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(15543)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17225

Query Match          76.4%; Score 16.8; DB 3; Length 15543;
Best Local Similarity 90.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 ACCTTCAACTCTGCGTCCTT 22
Db 4395 ACCTTCATCTCTGTGTCTT 4414

RESULT 8
US-08-684-862-10/c
; Sequence 10, Application US/08684862
; Patent No. 5759541
; GENERAL INFORMATION:
; APPLICANT: Bach, Alfred
; APPLICANT: Hillen, Heinz
; APPLICANT: Bialojan, Siegfried
; TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
; FILE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 5.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/684,862
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,705
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,040
; FILING DATE: 30-DEC-1992
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 5.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/684,862
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,705
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,040
; FILING DATE: 30-DEC-1992
; APPLICATION NUMBER: PCT/EP91/01361
; FILING DATE: 19-JUL-1991
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 988 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Agkistrodon rhodostoma
; FEATURE:
; LOCATION: 197 to 904
; OTHER INFORMATION: the coding region shown in (2)(ix)(B)
; OTHER INFORMATION: codes for the protein of SEQ ID NO: 5
US-08-684-862-10

Query Match          74.5%; Score 16.4; DB 2; Length 988;
Best Local Similarity 94.4%; Pred. No. 2.5e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATACCTTCAACTCTGCGT 18
Db 126 ATAGCTTCAACTCTGCGT 109

RESULT 9
US-08-684-862-9/c
; Sequence 9, Application US/08684862
; Patent No. 5759541
; GENERAL INFORMATION:
; APPLICANT: Bach, Alfred
; APPLICANT: Hillen, Heinz
; APPLICANT: Bialojan, Siegfried
; TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use
; FILE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage
; COMPUTER: IBM AT-compatible, 80286 processor
; OPERATING SYSTEM: MS-DOS version 5.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/684,862
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/361,705
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/966,040
; FILING DATE: 30-DEC-1992
; COMPUTER READABLE FORM:
```

```
; APPLICATION NUMBER: PCT/EP91/01361
; FILING DATE: 19-JUL-1991
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Agkistrodon rhodostoma
; FEATURE:
; LOCATION: 231 to 935
; OTHER INFORMATION: the coding region shown in (2) (ix) (B)
; OTHER INFORMATION: codes for the protein of SEQ ID NO: 4
US-08-684-862-9

Query Match 74.5%; Score 16.4; DB 2; Length 1333;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGT 18
||| ||||| ||||| |||||
Db 157 ATAGCTTCAACTCTGCGT 140

RESULT 10
US-09-949-016-71590
; Sequence 71590, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71590
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-71590

Query Match 73.6%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCT 21
||| ||||| ||||| |||||
Db 495 ATGCCTTGAACCTCGCATCCT 515

RESULT 11
US-09-949-016-179033
; Sequence 179033, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20

; APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179033
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-179033

Query Match 73.6%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCT 21
||| ||||| ||||| |||||
Db 556 ACACCTTCACCTCTGCTTCCT 576

RESULT 12
US-09-949-016-179034
; Sequence 179034, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 179034
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-179034

Query Match 73.6%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGTCCT 21
||| ||||| ||||| |||||
Db 116 ACACCTTCACCTCTGCTTCCT 136

RESULT 13
US-09-270-767-1498/c
; Sequence 1498, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1498
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-1498
```


Query Match 73.6%; Score 16.2; DB 3; Length 703;
 Best Local Similarity 85.7%; Pred. No. 2.9e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TACCTTCAACTCTGCGTCCTT 22
 Db 589 TGCCTTCAACTTGGCTCCAT 569

RESULT 14

US-09-270-767-16780/c
 ; Sequence 16780, Application US/09270767
 ; Patent No. 6703491
 ; GENERAL INFORMATION:
 ; APPLICANT: Homburger et al.
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
 ; FILE REFERENCE: File Reference: 7326-094
 ; CURRENT APPLICATION NUMBER: US/09/270,767
 ; CURRENT FILING DATE: 1999-03-17
 ; NUMBER OF SEQ ID NOS: 62517
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 16780
 ; LENGTH: 703
 ; TYPE: DNA
 ; ORGANISM: Drosophila melanogaster
 US-09-270-767-16780

Query Match 73.6%; Score 16.2; DB 3; Length 703;
 Best Local Similarity 85.7%; Pred. No. 2.9e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TACCTTCAACTCTGCGTCCTT 22
 Db 589 TGCCTTCAACTTGGCTCCAT 569

RESULT 15

US-09-976-594-1098/c
 ; Sequence 1098, Application US/09976594
 ; Patent No. 6673549
 ; GENERAL INFORMATION:
 ; APPLICANT: Furness, Michael
 ; APPLICANT: Buchbinder, Jenny
 ; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
 ; FILE REFERENCE: PA-0041 US
 ; CURRENT APPLICATION NUMBER: US/09/976,594
 ; CURRENT FILING DATE: 2001-10-12
 ; PRIOR APPLICATION NUMBER: 60/240,409
 ; PRIOR FILING DATE: 2000-10-12
 ; NUMBER OF SEQ ID NOS: 1143
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 1098
 ; LENGTH: 3698
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: Incyte ID No. 6673549 197159.1
 US-09-976-594-1098

Query Match 73.6%; Score 16.2; DB 3; Length 3698;
 Best Local Similarity 85.7%; Pred. No. 3.7e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TACCTTCAACTCTGCGTCCTT 22
 Db 2577 TTCTTCAAAATCTGTCTT 2557

Search completed: November 27, 2005, 01:10:52
 Job time : 6.556 secs

1715 Page Blank (uspro)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 26, 2005, 13:00:41 ; Search time 25,3623 Seconds
(without alignments)
7173.088 Million cell updates/sec

Title: US-09-555-529-28

Perfect score: 22

Sequence: 1 ataccttcaactctgcgtcctt 22

Scoring table: IDENTITY_NUC

Gapop 10,10 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.Main:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20.4	92.7	591	5	US-10-106-698-1187
C 2	18.8	85.5	25	7	US-10-719-956-48210
C 3	18.8	85.5	12332	7	US-10-236-542D-1
C 4	17.8	80.9	1432	7	US-10-425-114-32243
C 5	17.8	80.9	1921	7	US-10-425-114-32123
C 6	17.8	80.9	1995	8	US-10-425-115-167414
C 7	17.4	79.1	4301	7	US-10-437-963-11677
C 8	17.4	79.1	11673	10	US-11-097-143-28096
C 9	17.4	79.1	11707	10	US-11-097-143-36907
C 10	17.2	78.2	25	7	US-10-719-956-48209
C 11	17.2	78.2	637	5	US-10-027-632-226571
C 12	17.2	78.2	637	5	US-10-027-632-226572
C 13	17.2	78.2	637	5	US-10-027-632-226573
C 14	17.2	78.2	637	6	US-10-027-632-226571
C 15	17.2	78.2	637	6	US-10-027-632-226572
C 16	17.2	78.2	637	6	US-10-027-632-226573
C 17	17.2	78.2	825	7	US-10-282-122A-17789
C 18	17.2	78.2	2000	7	US-10-380-727-35
C 19	17.2	78.2	3814	5	US-10-102-524-1701
C 20	17.2	78.2	3876	5	US-10-034-934-50
C 21	17.2	78.2	111206	5	US-10-087-192-1957
C 22	17.2	78.2	367378	6	US-10-312-841-1
C 23	17.2	77.3	1858	7	US-10-425-114-15078

24	17	77.3	1961	7	US-10-425-114-29972	Sequence 29972, A
25	17	77.3	1974	7	US-10-425-114-29922	Sequence 29922, A
26	17	77.3	2617	7	US-10-424-599-101374	Sequence 101374, A
C 27	16.8	76.4	498	7	US-10-621-901-33	Sequence 33, Appl
C 28	16.8	76.4	502	6	US-10-062-674-716	Sequence 716, App
C 29	16.8	76.4	546	7	US-10-621-901-2071	Sequence 2071, App
C 30	16.8	76.4	648	5	US-10-027-632-276528	Sequence 276528, A
C 31	16.8	76.4	648	6	US-10-027-632-276528	Sequence 13619, A
C 32	16.8	76.4	715	5	US-10-027-632-13619	Sequence 13620, A
C 33	16.8	76.4	715	5	US-10-027-632-13620	Sequence 13621, A
C 34	16.8	76.4	715	5	US-10-027-632-13621	Sequence 13619, A
C 35	16.8	76.4	715	6	US-10-027-632-13619	Sequence 13620, A
C 36	16.8	76.4	715	6	US-10-027-632-13620	Sequence 13621, A
C 37	16.8	76.4	715	6	US-10-027-632-13621	Sequence 13621, A
C 38	16.8	76.4	715	6	US-10-425-114-30941	Sequence 30941, A
C 39	16.8	76.4	1234	8	US-10-425-115-110431	Sequence 110431, A
C 40	16.8	76.4	1266	8	US-10-156-761-4325	Sequence 110433, A
C 41	16.8	76.4	1542	6	US-10-156-761-4325	Sequence 4325, App
C 42	16.8	76.4	1589	3	US-09-866-050A-571	Sequence 571, App
C 43	16.8	76.4	1589	5	US-10-152-661-571	Sequence 40577, A
C 44	16.8	76.4	3472	7	US-10-437-963-40577	Sequence 37, Appl
C 45	16.8	76.4	13017	5	US-10-139-833-37	

ALIGNMENTS

RESULT 1
US-10-106-698-1187/c
; Sequence 1187, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1187
; LENGTH: 591
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (587)..(587)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1187

Query Match 92.7% Score 20.4; DB 5; Length 591;
Best Local Similarity 95.5%; Pred. No. 6.5;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 ATACCTTCAACTCTGCGTCCTT 22
|||
Db 239 ATTCCTTCAACTCTGCGTCCTT 218
|||

RESULT 2
US-10-719-956-48210/c
; Sequence 48210, Application US/10719956
; Publication No. US20040146910A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei
; TITLE OF INVENTION: Methods of Genetic Analysis of Rat
; FILE REFERENCE: 3527.1
; CURRENT APPLICATION NUMBER: US/10/719,956

Db 1628 TTCCTTCAGCTCTGCGTCCTT 1608

RESULT 7

US-10-437-963-11677
; Sequence 11677, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 11677
; LENGTH: 4301
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_17879C.1
US-10-437-963-11677

Query Match 79.1%; Score 17.4; DB 7; Length 4301;
Best Local Similarity 94.7%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 ACCTTCAACTCTGCGTCCT 21
Db 763 ACCTTCAACTCTGCGTCCT 781

RESULT 8

US-11-097-143-28096/c
; Sequence 28096, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28096
; LENGTH: 11673
; TYPE: DNA

; ORGANISM: DROSOPHILA
US-11-097-143-28096

Query Match 79.1%; Score 17.4; DB 10; Length 11673;
Best Local Similarity 94.7%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CCTTCAACTCTGCGTCCTT 22
Db 4619 CCTTCAACTCTGCGTCCTT 4601

RESULT 9

US-11-097-143-36907/c
; Sequence 36907, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36907
; LENGTH: 11707
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-36907

Query Match 79.1%; Score 17.4; DB 10; Length 11707;
Best Local Similarity 94.7%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CCTTCAACTCTGCGTCCTT 22
Db 4619 CCTTCAACTCTGCGTCCTT 4601

RESULT 10

US-10-719-956-48209/c
; Sequence 48209, Application US/10719956
; Publication No. US20040146910A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Rat
; FILE REFERENCE: 3527.1
; CURRENT APPLICATION NUMBER: US/10/719,956
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,836
; PRIOR FILING DATE: 2002-11-20
; NUMBER OF SEQ ID NOS: 699466
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 48209

```
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-10-719-956-48209

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 7; Length 25;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGGTCCTT 22
Db 22 ATACCTTCAACTCTCGGTCCTT 1

RESULT 11
US-10-027-632-226571
; Sequence 226571, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 226571
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226571

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 5; Length 637;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGGTCCTT 22
Db 410 ACACCTTCACTCTCGGTCCTT 431

RESULT 12
US-10-027-632-226572
; Sequence 226572, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 226572
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226572

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 5; Length 637;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGGTCCTT 22
Db 410 ACACCTTCACTCTCGGTCCTT 431

RESULT 13
US-10-027-632-226573
; Sequence 226573, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 226573
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226573

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 5; Length 637;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTCGGTCCTT 22
Db 410 ACACCTTCACTCTCGGTCCTT 431

RESULT 14
US-10-027-632-226571
; Sequence 226571, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
```


This Page Blank (uspio)


```
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 813058
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-813058
```

```
Query Match 79.1%; Score 17.4; DB 9; Length 19;
Best Local Similarity 94.7%; Pred. No. 3.4;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TACCTTCAACTCTGCGTCC 20
Db 19 TTCTTCAACTCTGCGTCC 1
```

```
RESULT 3
US-11-101-244-813016/c
; Sequence 813016, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmakon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 813016
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-813016
```

```
Query Match 74.5%; Score 16.4; DB 8; Length 19;
Best Local Similarity 94.4%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGT 18
Db 18 ATTCTTCAACTCTGCGT 1
```

```
RESULT 4
US-11-083-784-813016/c
; Sequence 813016, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmakon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
```

```
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 813016
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-813016
```

```
Query Match 74.5%; Score 16.4; DB 9; Length 19;
Best Local Similarity 94.4%; Pred. No. 10;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATACCTTCAACTCTGCGT 18
Db 18 ATTCTTCAACTCTGCGT 1
```

```
RESULT 5
US-10-858-730-150/c
; Sequence 150, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Thermobifida fusca
US-10-858-730-150
```

```
Query Match 69.1%; Score 15.2; DB 1; Length 1290;
Best Local Similarity 85.0%; Pred. No. 59;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ACCTTCAACTCTGCGTCTT 22
Db 857 ACCTTCAACGCTACGCGCTT 838
```

```
RESULT 6
US-11-102-978-3
; Sequence 3, Application US/11102978
```

Publication No. US20050250142A1
GENERAL INFORMATION:
APPLICANT: University of Utah Technology Transfer Office
APPLICANT: University of Utah Research Foundation
TITLE OF INVENTION: Diagnosis and Treatment of Herpes Simplex Virus Disease
FILE REFERENCE: 0274-5537.1US
CURRENT APPLICATION NUMBER: US/11/102,978
CURRENT FILING DATE: 2005-04-11
PRIOR APPLICATION NUMBER: PCT/US2003/033152
PRIOR FILING DATE: 2003-10-18
PRIOR APPLICATION NUMBER: 60/419,576
PRIOR FILING DATE: 2002-10-18
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn version 3.2
SEQ ID NO 3
LENGTH: 340000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: exon
LOCATION: (56948)..(57115)
OTHER INFORMATION: C21orf34 exon
FEATURE:
NAME/KEY: misc feature
LOCATION: (80066)..(81089)
OTHER INFORMATION: Gene VDACC2P; voltage-dependent anion channel isoform 2 pseudogene
FEATURE:
NAME/KEY: exon
LOCATION: (167308)..(167438)
OTHER INFORMATION: C21orf34 exon
FEATURE:
NAME/KEY: exon
LOCATION: (216732)..(216833)
OTHER INFORMATION: C21orf34 exon
US-11-102-978-3

Query Match 69.1%; Score 15.2; DB 7; Length 340000;
Best Local Similarity 85.0%; Pred. No. 90;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2 TACCTCAACTCTGCGTCCT 21
Db 263712 TACATTCACCTTTCCTCCT 263731

RESULT 7
US-11-101-244-1543865
Sequence 1543865, Application US/11/101244
Publication No. US20050246794A1
GENERAL INFORMATION:
APPLICANT: Dharmacon, Inc.
APPLICANT: Khvorova, Anastasia
APPLICANT: Reynolds, Angela
APPLICANT: Leake, Devin
APPLICANT: Marshall, William
APPLICANT: Scaringe, Stephen
TITLE OF INVENTION: Functional and Hyperfunctional siRNA
FILE REFERENCE: 13499US
CURRENT APPLICATION NUMBER: US/11/101,244
CURRENT FILING DATE: 2005-04-07
PRIOR APPLICATION NUMBER: 60/502,050
PRIOR FILING DATE: 2003-09-10
PRIOR APPLICATION NUMBER: 60/426,137
PRIOR FILING DATE: 2002-11-14
NUMBER OF SEQ ID NOS: 1591911
SOFTWARE: Proprietary
SEQ ID NO 1543865
LENGTH: 19
TYPE: RNA
ORGANISM: Homo sapiens
US-11-101-244-1543865

Query Match 67.3%; Score 14.8; DB 8; Length 19;
Best Local Similarity 81.0%; Pred. No. 1.2e+02;

Best Local Similarity 50.0%; Pred. No. 62;
Matches 9; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
Qy 5 CTTCACCTCTGCGTCCTT 22
Db 1 CUUUAUAAUCUGCGUCCUU 18
RESULT 8
US-11-083-784-1543865
Sequence 1543865, Application US/11/083784
Publication No. US20050245475A1
GENERAL INFORMATION:
APPLICANT: Dharmacon, Inc.
APPLICANT: Khvorova, Anastasia
APPLICANT: Reynolds, Angela
APPLICANT: Leake, Devin
APPLICANT: Marshall, William
APPLICANT: Scaringe, Stephen
TITLE OF INVENTION: Functional and Hyperfunctional siRNA
FILE REFERENCE: 13499US
CURRENT APPLICATION NUMBER: US/11/083,784
CURRENT FILING DATE: 2005-03-18
PRIOR APPLICATION NUMBER: US/10/714,333
PRIOR FILING DATE: 2003-11-14
PRIOR APPLICATION NUMBER: 60/502,050
PRIOR FILING DATE: 2003-09-10
PRIOR APPLICATION NUMBER: 60/426,137
PRIOR FILING DATE: 2002-11-14
NUMBER OF SEQ ID NOS: 1591911
SOFTWARE: Proprietary
SEQ ID NO 1543865
LENGTH: 19
TYPE: RNA
ORGANISM: Homo sapiens
US-11-083-784-1543865

Query Match 67.3%; Score 14.8; DB 9; Length 19;
Best Local Similarity 50.0%; Pred. No. 62;
Matches 9; Conservative 7; Mismatches 2; Indels 0; Gaps 0;
Qy 5 CTTCACCTCTGCGTCCTT 22
Db 1 CUUUAUAAUCUGCGUCCUU 18

RESULT 9
US-11-012-762-67
Sequence 67, Application US/11/012762
Publication No. US20050244815A1
GENERAL INFORMATION:
APPLICANT: Georgia State University Research Foundation, Inc.
TITLE OF INVENTION: Compositions and Methods for Viral Resistance Genes
FILE REFERENCE: GSU1.PCT
CURRENT APPLICATION NUMBER: US/11/012,762
CURRENT FILING DATE: 2004-12-15
PRIOR APPLICATION NUMBER: PCT/US03/19300
PRIOR FILING DATE: 2003-06-19
PRIOR APPLICATION NUMBER: US 60/390,046
PRIOR FILING DATE: 2002-06-19
NUMBER OF SEQ ID NOS: 133
SOFTWARE: PatentIn version 3.2
SEQ ID NO 67
LENGTH: 2311
TYPE: DNA
ORGANISM: Rattus norvegicus
FEATURE:
NAME/KEY: CDS
LOCATION: (41)..(2242)
US-11-012-762-67

Query Match 66.4%; Score 14.6; DB 9; Length 2311;
Best Local Similarity 81.0%; Pred. No. 1.2e+02;

Matches	17;	Conservative	0;	Mismatches	4;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

QY 1 ATACCTTCAACTCTGCGTCT 21
1933 ATACCGGAGCTCTGCGTCT 1953

```

RESULT 10
US-11-101-244-495567/c
; Sequence 495567, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13498US
; CURRENT APPLICATION NUMBER: US/11/101.244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 495567
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-495567

```

```

RESULT 11
US-11-083-784-495567/c
; Sequence 495567, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 495567
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-495567
Query Match 64.5%; Score 14.2; DB 9; Length

```

Best Local Similarity 84.2%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RESULT 12

US-10-995-793-72/c
; Sequence 72, Application US/10995793
; Publication No. US20050250123A1
; GENERAL INFORMATION:
; APPLICANT: Yang, Ri-Yao
; APPLICANT: Hsu, Daniel K.
; APPLICANT: Liu, Fu-Tong
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Reducing Galectin-12 Activity to Reduce Formation of
; TITLE OF INVENTION: Adipocytes
; FILE REFERENCE: 023070-139910US
; CURRENT APPLICATION NUMBER: US/10/995,793
; CURRENT FILING DATE: 2004-11-22
; PRIOR APPLICATION NUMBER: US 60/524,418
; PRIOR FILING DATE: 2003-11-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 72
; LENGTH: 2468
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse galectin-12 CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (141)..(1085)
; OTHER INFORMATION: mouse galectin-12
; FEATURE:
; NAME/KEY: repeat region
; LOCATION: (1233)..(1282)
; OTHER INFORMATION: CA repeats
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2336)..(2340)
; OTHER INFORMATION: AT-rich motif in 3'-UTR
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2350)..(2354)
; OTHER INFORMATION: AT-rich motif in 3'-UTR
US-10-995-793-72

Query Match 64.5%; Score 14.2; DB 1; Length 2468;
Best Local Similarity 84.2%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps

QY 3 ACCTTCAACTTCGCGTCT 21
||||| ||||| |||||
Db 514 ACCTTCACTCTTCGTCT 496

RESULT 13

US-11-113-837-3
; Sequence 3, Application US/11113837
; Publication No. US2005025513A1
; GENERAL INFORMATION:
; APPLICANT: Akin, Ali R.
; APPLICANT: Bodie, Elizabeth A.
; APPLICANT: Burrow, Shirley
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Turner, Geoffrey
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Regulatable Growth of Filamentous Fungi
; FILE REFERENCE: GC682-2
CURRENT APPLICATION NUMBER: US/11/113,837

```

RESULT 13
US-11-113-837-3
; Sequence 3, Application US/11113837
; Publication No. US20050255513A1
; GENERAL INFORMATION:
; APPLICANT: Akin, Ali R.
; APPLICANT: Bodie, Elizabeth A.
; APPLICANT: Burrow, Shirley
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Turner, Geoffrey
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Regulatable Growth of Filamentous Fungi
; FILE REFERENCES: GC682-2
; CURRENT APPLICATION NUMBER: US/11/113,837

```

CURRENT FILING DATE: 2005-04-25
PRIOR APPLICATION NUMBER: US 60/276,571
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: US 60/276,618
PRIOR FILING DATE: 2001-03-14
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 3038
TYPE: DNA
ORGANISM: Aspergillus nidulans
US-11-113-837-3

Query Match 64.5%; Score 14.2; DB 7; Length 3038;
Best Local Similarity 84.2%; Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TACCTTCAACTCTCGTCC 20
Db 780 TGCCTTCAACCTCGGCC 798

RESULT 14
US-11-113-837-14
Sequence 14, Application US/11113837
Publication No. US20050255513A1
GENERAL INFORMATION:
APPLICANT: Akin, Ali R.
APPLICANT: Bodie, Elizabeth A.
APPLICANT: Burrow, Shirley
APPLICANT: Dunn-Coleman, Nigel
APPLICANT: Turner, Geoffrey
APPLICANT: Ward, Michael
TITLE OF INVENTION: Regulatable Growth of Filamentous Fungi
FILE REFERENCE: GC682-2
CURRENT APPLICATION NUMBER: US/11/113,837
CURRENT FILING DATE: 2005-04-25
PRIOR APPLICATION NUMBER: US 60/276,571
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: US 60/276,618
PRIOR FILING DATE: 2001-03-14
NUMBER OF SEQ ID NOS: 21
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 3038
TYPE: DNA
ORGANISM: Aspergillus nidulans
US-11-113-837-14

Query Match 64.5%; Score 14.2; DB 7; Length 3038;
Best Local Similarity 84.2%; Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 TACCTTCAACTCTCGTCC 20
Db 780 TGCCTTCAACCTCGGCC 798

RESULT 15
US-10-131-826A-541
Sequence 541, Application US/10131826A
Publication No. US20050245730A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.

APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C128
CURRENT APPLICATION NUMBER: US/10/131,826A
CURRENT FILING DATE: 2002-04-24
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-19
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 541
LENGTH: 3824
TYPE: DNA
ORGANISM: Homo Sapien
US-10-131-826A-541

Query Match 64.5%; Score 14.2; DB 1; Length 3824;
Best Local Similarity 84.2%; Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 CCTTCAACTCTCGTCCCTT 22
Db 1003 CCTTCAACTCTGTGCCCTT 1021

Search completed: November 27, 2005, 02:44:32
Job time : 8.76471 secs

2008 0000 0000